

DEVICE TECHNOLOGY DEVELOPMENT FOR  
ULTRA-LOW-POWER CIRCUIT DESIGN

A DISSERTATION  
SUBMITTED TO THE DEPARTMENT OF ELECTRICAL ENGINEERING  
AND THE COMMITTEE ON GRADUATE STUDIES  
OF STANFORD UNIVERSITY  
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS  
FOR THE DEGREE OF  
DOCTOR OF PHILOSOPHY

Jawad Nasrullah

March 2005

UMI Number: 3162308

Copyright 2005 by  
Nasrullah, Jawad

All rights reserved.

#### INFORMATION TO USERS

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleed-through, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.



---

UMI Microform 3162308

Copyright 2005 by ProQuest Information and Learning Company.  
All rights reserved. This microform edition is protected against  
unauthorized copying under Title 17, United States Code.

ProQuest Information and Learning Company  
300 North Zeeb Road  
P.O. Box 1346  
Ann Arbor, MI 48106-1346

© Copyright by Jawad Nasrullah 2005  
All Rights Reserved

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

G. Leonard Tyler  
G. Leonard Tyler  
(Principal Adviser)

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

Yoshio Nishi  
Yoshio Nishi

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

Howard A. Zebker  
Howard A. Zebker

Approved for the University Committee on Graduate Studies.

*To my 2 sweetie  $\pi$ 's, Iram and Mariam*

# Abstract

Ultra-low-power (ULP) circuit techniques incorporating dynamic threshold-voltage control through back-gate bias have promise for power control of leading-edge fine-geometry CMOS. The use of an edge-defined patterning process makes it possible to aggressively reduce the scale of thin-film transistors with back-gate-bias threshold-voltage control in a 3-D compatible IC process. With this approach, nanometer (nm) scale lines patterned in silicon, using 1- $\mu$ m optical lithography with standard materials and standard processing equipment, are compatible with low-thermal-budget processes. For implementation, chemical-vapor deposition (CVD) defines spacers around optically-registered edges. These spacers are combined subsequently with a photoresist mask to pattern underlying layers. Ability to control variations in CVD and dry etching processes to less than 5% enables tight spacer-width control for patterning of features in the range of 18 to 180 nm. Local critical-dimension (CD) variations of up to 7 nm, based on CD-SEM analysis, are observed in test structures. Spacer-deposition roughness, oxide-etch chemistry, and edge-registration lithography contribute to CD variations and line-edge roughness. Analysis of deposited silicon and silicon-dioxide film surface roughness demonstrates that films thinner than about 20 nm are well formed when deposited more thickly than needed, and then reduced by etching to the desired dimension. Use of a binary process, consisting of alternating etching in hydrofluoric acid and nitric acid, improves roughness control. Insulated back-gated prototype thin-film transistors incorporate a 90-nm gate-length created by the use of edge-defined lithographic process. Threshold voltages of these devices are adjustable by about 100 mV per volt of back-gate bias. Intra-die variations are

observed in the electrical characteristics of these devices. It is hypothesized that these variations are the result of polycrystalline-silicon grain-boundary charge traps, and channel shortening due to enhanced dopant diffusion along grain boundaries.

# Acknowledgments

Jawad Nasrullah  
jawad@stanfordalumni.org

*Stanford, California*  
*March 16, 2005*

Completion of this work was made possible through Professor G. Leonard Tyler's discretionary funds with generous gift contributions by SUN Microsystems. The Department of Electrical Engineering and the Stanford Center for Professional Development at Stanford University also supported me at times. Some support for scanning electron microscope studies was provided by Sun Microsystems and by Hitachi High-Technologies America. NanoDevices Inc. generously provided materials for AFM scanning.

I thank James Burr of Transmeta Corporation for being my mentor and providing motivation for this work. I am also grateful to Al Matthews of Agilent Technologies for guiding me through the work on film surfaces and device integration. Dr. Eric Perozziello at SNF taught me many practical techniques of semiconductor processing, and I thank him for that. Dr. Kleanthes Koniaris of Transmeta Corporation reviewed this dissertation in detail and provided helpful feedback. I greatly appreciate the editorial support provided by Chuck Hastings of Federal Way, Washington. I also acknowledge help from Chung Lam of Sun Microsystems, Daniel Becker of Hitachi, and Hector Cavazos of NanoDevices.

I am extremely grateful to my Ph.D. adviser, Professor G. Leonard Tyler, for years of untiring guidance and support for this work, and for being a great teacher and a great example of scholarship. I am also extremely indebted to my associate

Ph.D. adviser, Professor Yoshio Nishi, for guidance, useful discussions, and helpful comments. Professor Howard Zebker very kindly agreed to serve as the third reader, and I am thankful to him for that.

A number of friends have been very helpful all along, including Vjekoslav Svilan, Husrev T. Ilhan, Bevan M. Baas, Karen Huyser, Amol Joshi, and Talal Alattar. The starting L<sup>A</sup>T<sub>E</sub>X framework for this thesis was acquired from the theses of former STAR Lab students Zartash Uzmi, Bilal Ahmad, and John Baron.

In the end I would like to acknowledge my family, especially the support I received from my grandfather Jalal-ud-Din Chughtai who passed away while I was away working on this dissertation. Last, but not least, this work was really made possible by the love, support, and understanding of my sweethearts Iram Mirza and Mariam Nasrullah.

*“I designed, fabricated, and tested these devices.”*

—Jawad Nasrullah

March 16, 2005

# Contents

<b>Abstract</b>	<b>v</b>
<b>Acknowledgments</b>	<b>vii</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Organization of the Dissertation . . . . .	1
1.2 Contributions . . . . .	2
1.3 Objectives of this Research . . . . .	3
<b>2 Overview of Low-Power Design</b>	<b>4</b>
2.1 History of Stanford’s Ultra-Low-Power (ULP) Technology . . . . .	6
2.2 Why Worry About Power Dissipation? . . . . .	7
2.2.1 Batteries . . . . .	8
2.2.2 Computing-circuitry density . . . . .	9
2.2.3 Environmental impact . . . . .	10
2.2.4 System considerations . . . . .	10
2.3 What’s Inside Those Power-Hungry Integrated Circuits? . . . . .	12
2.4 Power-Reduction Techniques for Circuits . . . . .	18
2.4.1 Parameters to consider for power optimization . . . . .	18
2.4.2 Power optimization circuit techniques . . . . .	21
<b>3 Device Requirements for ULP Technology</b>	<b>28</b>
3.1 Power Dissipation in CMOS . . . . .	28
3.1.1 Active power dissipation . . . . .	29

3.1.2	Short-circuit power dissipation . . . . .	30
3.1.3	Leakage power dissipation . . . . .	30
3.1.4	Drain-induced barrier lowering . . . . .	33
3.2	ULP Techniques . . . . .	34
3.3	Characteristics of ULP Devices . . . . .	35
3.3.1	Threshold-voltage tunability . . . . .	35
3.3.2	Turn-off behavior . . . . .	36
3.3.3	3-D device stacking . . . . .	37
<b>4</b>	<b>Thin-Film Surfaces for ULP Device Technology</b>	<b>38</b>
4.1	Overview of Thin-Film Deposition Processes . . . . .	40
4.1.1	Thermal growth of SiO <sub>2</sub> . . . . .	41
4.1.2	Low-pressure chemical vapor deposition . . . . .	41
4.1.3	Deposition by ‘evaporated’ Si . . . . .	42
4.2	Experimental Details and Results . . . . .	43
4.2.1	Sample preparation, scanning, and characterization . . . . .	43
4.2.2	Bare Si wafer surfaces . . . . .	45
4.2.3	SiO <sub>2</sub> substrate surfaces . . . . .	45
4.2.4	Thin Si film surfaces . . . . .	46
<b>5</b>	<b>Edge-Defined Lithography</b>	<b>58</b>
5.1	Lithography . . . . .	60
5.1.1	Lithographic techniques . . . . .	60
5.1.2	Lithography selection criterion . . . . .	63
5.2	Edge-Defined Lithography . . . . .	65
5.2.1	Processing details . . . . .	65
5.3	Experimental Results . . . . .	71
5.3.1	SEM images during edge-defined patterning . . . . .	71
5.3.2	Critical dimension and line-edge roughness . . . . .	74
5.3.3	Sample preparation . . . . .	74

5.3.4	Undesirable trench formation . . . . .	76
5.3.5	180-nm edge-defined patterning . . . . .	78
5.3.6	90-nm edge-defined patterning . . . . .	80
5.3.7	$\leq$ 50-nm edge-defined patterning . . . . .	81
5.3.8	CD-SEM analysis . . . . .	83
5.3.9	Design rules . . . . .	84
5.3.10	Defect tolerance . . . . .	88
5.4	Mask Layout Techniques for Edge-Defined Lithography . . . . .	90
5.4.1	Footprint reduction . . . . .	90
5.4.2	Multiple gates . . . . .	91
5.4.3	Serpentine edges . . . . .	91
5.4.4	Experimental limitations . . . . .	93
<b>6</b>	<b>Device and Process Design</b>	<b>94</b>
6.1	ULP Device Structure . . . . .	95
6.1.1	An ideal ULP device . . . . .	95
6.1.2	Design challenges . . . . .	96
6.1.3	The prototype device . . . . .	99
6.2	Device Optimization . . . . .	101
6.2.1	The device design rules . . . . .	101
6.2.2	$L_g = 40$ -nm device design sensitivity . . . . .	102
6.3	Device Processing Flow . . . . .	107
6.3.1	Process simulation . . . . .	110
6.3.2	Processing details . . . . .	111
6.4	The Completed Device . . . . .	118
<b>7</b>	<b>Device Results</b>	<b>124</b>
7.1	Sheet Resistances . . . . .	126
7.2	Device Curves . . . . .	128
7.3	Effect of Back-Gate Bias . . . . .	131

7.4	Temperature Dependence of Device Performance . . . . .	141
7.5	Intra-Die Device Performance Variation . . . . .	145
<b>8</b>	<b>Summary and Future Work</b>	<b>151</b>
8.1	Summary . . . . .	152
8.2	Contributions . . . . .	154
8.3	Future Work and Improvements . . . . .	155
8.3.1	Materials . . . . .	155
8.3.2	Edge-Defined Lithography . . . . .	156
8.3.3	ULP device optimization . . . . .	157
8.4	Concluding Remarks . . . . .	157
<b>Appendices</b>		<b>158</b>
<b>A</b>	<b>Relevant Device Physics</b>	<b>159</b>
A.1	Mass Action Law . . . . .	159
A.2	Fermi Energy . . . . .	161
A.3	Poisson's Equation . . . . .	162
A.4	Threshold Voltage for a Bulk MOS Transistor . . . . .	162
A.5	Threshold Voltage of an Asymmetric Double-Gate TFT . . . . .	164
A.6	Drain Current of a Long-Channel Bulk MOS Transistor . . . . .	165
A.7	Drain Current of a Short-Channel Bulk MOS Transistor . . . . .	165
A.8	Off-Current of an MOS Transistor . . . . .	166
A.9	Power Dissipation in CMOS . . . . .	166
<b>B</b>	<b>Processing Details for SNF</b>	<b>167</b>
B.1	Cleaning Procedures . . . . .	169
B.2	Wafer Scribing . . . . .	170
B.3	Gate-Oxide and Gate-Poly-Si Deposition . . . . .	171
B.4	Stop-Oxide and Registration Poly-Si Deposition . . . . .	172
B.5	Registration Photolithography . . . . .	173

B.6	Registration Poly-Si Patterning . . . . .	174
B.7	Spacer Deposition . . . . .	174
B.8	Oxide-Spacer Etching . . . . .	175
B.9	Registration Poly-Si Stripping . . . . .	175
B.10	Oxide-Mask Formation Etching . . . . .	175
B.11	Gate-formation etching . . . . .	175
<b>References</b>		<b>177</b>

# List of Tables

2.1	Energy content of batteries . . . . .	8
6.1	Optimal dimensions of a TFT with gate length=40 nm . . . . .	96
6.2	Dimensions of prototype TFT with gate length=90 nm . . . . .	100
6.3	Device dimensions used for sensitivity analysis . . . . .	102
6.4	Device dimensions used for sensitivity analysis . . . . .	115
7.1	Sheet resistivity measurements after 1-sec RTA . . . . .	126
7.2	Sheet resistivity measurements after 3-sec RTA . . . . .	127
7.3	Contact-chain resistance . . . . .	128
7.4	Test structures . . . . .	129
A.1	Constants . . . . .	159
A.2	Variables . . . . .	160
A.3	Silicon dopant impurities . . . . .	161
B.1	Pieces of equipment used at SNF . . . . .	168
B.2	Processes used at SNF . . . . .	169
B.3	Photoresist-stripping procedure . . . . .	169
B.4	Pre-LPCVD cleaning procedure . . . . .	170
B.5	Lampoly poly-Si etching recipe . . . . .	174

# List of Figures

2.1	Commercially available heatsinks for Intel microprocessors . . . . .	11
2.2	Example of a microprocessor . . . . .	12
2.3	Typical datapath for synchronous circuit design . . . . .	14
2.4	Switching activity through a typical datapath circuit . . . . .	15
2.5	Static CMOS inverter . . . . .	16
2.6	A static-CMOS-inverter chain . . . . .	17
2.7	Currents through a switching inverter . . . . .	17
2.8	Application space for power-reduction circuit techniques . . . . .	22
2.9	Supply gating . . . . .	23
2.10	A CMOS inverter with separated back-gate terminals . . . . .	26
3.1	Leakage currents in a MOS transistor . . . . .	31
3.2	Typical turn-off characteristics of a MOS transistor . . . . .	32
3.3	Typical turn-off behavior for different drain voltages . . . . .	33
3.4	Typical $I_D$ - $V_D$ characteristics of a MOS transistor . . . . .	34
4.1	AFM scans of film surfaces . . . . .	47
4.2	Progression of surface-film growth . . . . .	49
4.3	Surface scans of 9-, 18-, and 36-nm thick Si films on thermal oxide .	50
4.4	Surface scans of 9-, 18-, and 36-nm thick Si films on LPCVD $\text{SiO}_2$ .	51
4.5	$\Delta h_{\text{p-p}}$ of Si films deposited on $\text{SiO}_2$ substrate . . . . .	53
4.6	Effect of predeposition cleaning . . . . .	54
4.7	Effect of etching back . . . . .	55

4.8	Effect of ion implantation and crystallization . . . . .	56
4.9	Surfaces of Si films deposited by evaporation of an Si source . . . . .	57
5.1	Example improvements from OPC . . . . .	61
5.2	Photoresist ashing . . . . .	63
5.3	Edge-defined patterning process flow . . . . .	67
5.4	Edge-defined patterning-spacer mask removal process . . . . .	72
5.5	SEM images for an edge-defined patterning experiment . . . . .	73
5.6	Definitions of CD and LER . . . . .	75
5.7	Trenching problem in the edge-defined patterning flow . . . . .	77
5.8	180-nm edge-defined pattern with an undesired trench . . . . .	78
5.9	180-nm edge-defined lithography results . . . . .	79
5.10	90-nm edge-defined patterned structure . . . . .	81
5.11	50-nm edge-defined patterned structure . . . . .	82
5.12	Side view of a 50-nm structure with oxide spacer mask . . . . .	82
5.13	25-nm structure without oxide mask . . . . .	83
5.14	18-nm edge-defined structure . . . . .	83
5.15	CD-SEM analysis of edge-defined structures . . . . .	85
5.16	Defining film-stack thicknesses for an edge-defined process . . . . .	89
5.17	Defect tolerance of the edge-definition process . . . . .	89
5.18	A single edge-defined gate finger . . . . .	90
5.19	Three edge-defined transistor fingers . . . . .	91
5.20	Two edge-defined transistors fabricated with one registration polygon	92
5.21	An edge-defined serpentine gate structure . . . . .	92
6.1	An ideal thin film transistor device . . . . .	96
6.2	TFT structure fabricated in this work . . . . .	100
6.3	$V_T$ variations vs. device dimensions . . . . .	104
6.4	Subthreshold slope vs. device dimensions . . . . .	105
6.5	On-current vs. device dimensions . . . . .	106

6.6	Off-current vs. device dimensions . . . . .	107
6.7	Edge-defined transistor fabrication . . . . .	109
6.8	Simulated steps in edge-defined transistor fabrication . . . . .	112
6.9	Alignment verniers . . . . .	115
6.10	Metal etching . . . . .	117
6.11	Drawn footprint of a triple-front-gate EdgeFET . . . . .	118
6.12	A prototype triple-front-gate EdgeFET . . . . .	119
6.13	AFM scan of a dual-front-gate EdgeFET . . . . .	120
7.1	Contact chain structure . . . . .	127
7.2	Device curves for a device with $L_g=90$ nm and $W_g=2000$ nm . . . . .	130
7.3	Device curves for a device with $L_g=360$ nm and $W_g=3000$ nm . . . . .	132
7.4	Device curves after Watts and Lee (1993) . . . . .	133
7.5	Effect of back-gate bias ( $V_B$ ) for $L_g=90$ nm, $W_g=1500$ nm . . . . .	135
7.6	Effect of back-gate bias ( $V_B$ ) for $L_g=180$ nm, $W_g=2000$ nm . . . . .	138
7.7	Effect of temperature on a device with $L_g=90$ nm, $W_g=2000$ nm . . . . .	142
7.8	Intra-die variations for a device with $L_g=90$ nm, $W_g=2000$ nm . . . . .	147
7.9	Poly-Si grain structure . . . . .	148
7.10	Effect of grain boundaries on TFT performance . . . . .	150

*This thesis breaks new ground in the fabrication and characterization of power-friendly integrated circuits having very small feature sizes, but produced using established procedures and standard equipment. Many of its results are applicable to on-chip 3-D vertical stacking of active devices.*

# Chapter 1

## Introduction

### 1.1 Organization of the Dissertation

This thesis describes the development of device-technology components for ultra-low-power (ULP) applications. ULP devices must have short channels and adjustable characteristics, and must be amenable to on-chip 3-D vertical stacking.

The final objective of this work is fabrication of transistors with gate lengths below 90 nm. These transistors must have dynamically adjustable threshold voltages controllable by back-gate biasing. Even though they are subject to 3-D vertical stacking, they are fabricated with standard semiconductor fabrication materials and processing equipment, in a manner that lends itself to high-throughput manufacturing.

Transistor fabrication techniques require deposition of thin films. The surface roughness of these films greatly affects the devices' characteristics. An atomic force microscope (AFM) is used to characterize the surfaces of silicon and silicon dioxide films. By employing a novel edge-defined nanolithography technique, we have created features as small as 18 nm using lab equipment otherwise considered capable only of 1000-nm feature resolution. Edge-defined patterning enables implementing thin-film transistors that show about 100 mV/V threshold voltage adjustability for applied back-gate bias.

This thesis comprises eight chapters, including this short introduction which describes the organization and plan of the dissertation and outlines the contributions. In Chapter 2 we describe the background of this work, and the significance of addressing power-dissipation issues in integrated circuits. We then present a survey of low-power circuit-design techniques and their applications. In Chapter 3 we focus on device aspects of the ultra-low-power technology that need attention, and we describe the requirements for ultra-low-power devices. In Chapter 4 we present characterization results for nanometer-scale roughness of the thin films of silicon and silicon dioxide that are to be used subsequently. In Chapter 5 we present results for edge-defined nanolithography. We show the experimental results of patterning feature sizes for 180, 90, 50, 25, and 18 nm. We also give details of the CD variation measurements, as well as advanced lithography approaches. In Chapter 6 we cover the design of the devices to be fabricated, and the process design that uses edge-defined lithography to fabricate thin-film transistors with insulated back-gates. In Chapter 7 we present the results of characterizing these prototype devices. Chapter 8 presents a summary of the dissertation, and suggestions for future research work.

## 1.2 Contributions

Contribution of this dissertation to the field of low-power electronics are listed below:

- Development of a new process for realizing electrically adjustable nanometer-scale transistors, and the creation of critical process modules.
- Characterization of deposited film stacks for obtaining smooth films, and development of an amorphous silicon binary etch-back smoothing process.
- Implementation of a low-temperature, edge-defined, nanometer-scale patterning technique using standard lithography equipment, compatible with existing CMOS processes.
- Novel edge-defined layout techniques for reducing mask complexity.
- Evaluation and optimization of device design.

- Fabrication of back-gated, solid-phase, crystallized thin-film transistors with adjustable characteristics.

### 1.3 Objectives of this Research

The objectives of this research include the design and development of a device technology that supports ULP circuit techniques; in particular, a technology that offers

- A method for transistor gate-length reduction to a few nanometers.
- Understanding of the challenges for nanometer-scale processes.
- Fabrication of transistor devices with adjustable  $V_T$ .

In achieving these objectives, the following restrictions were observed:

- All the processing is compatible with 3-D device stacking by controlling the temperature exposure carefully.
- All the processing is standard in terms of conventional CMOS processing flow.
- We limit processing to the use of conventional semiconductor tools, equipment, and materials.
- The process is compatible with lateral crystallization techniques.

*We now begin by surveying some of the work already done in the field of ultra-low-power design. A substantial portion of this work has been performed within the Stanford community, by faculty, students, and graduates. We will build upon this work in subsequent chapters.*

## Chapter 2

### Overview of Low-Power Design

This chapter presents a brief history of Stanford’s ultra-low-power (ULP) technology development. In order to understand the need for low-power design, we then discuss IC power dissipation in relation to battery life, computational-circuit density, the environment, and system design. Next, we look at the constituents of a typical IC, and its power dissipation in CMOS technologies, and some of the many ways in which researchers have developed, studied, and tested different circuit techniques in order to “push the envelope” for attaining better circuit performance at lower power-dissipation levels. Lastly, we present an overview of these different circuit techniques, including ULP, that have been proposed for power optimization.

During the 1980s, members of the Space, Telecommunication and Radio Science (STAR) laboratory of the Department of Electrical Engineering at Stanford University asked an important question: “Can a satellite be built that operates with just the solar energy falling on its surface?” The audience hearing this question included engineers, who set out to understand and optimize the energy efficiency of the electronics used in space applications. One outcome was the formation of the Ultra-Low-Power (ULP) research group, and the launching of the research program that followed. That question and its answer have become relevant to many electronics designs.

In the design and application of integrated circuits (ICs), power dissipation has always been a major constraint. For example, computer engineers face pressure to deliver designs with increasing computing power in ever-shrinking enclosures. Such

designs require efficient heat-removal methods and the use of power-efficient ICs. Traditionally, liquid and refrigerative cooling were used in earlier large computer systems, and heat-removal requirements dictated many design decisions. Due to their low cost, personal desktop computers have been fan-cooled. This type of cooling, however, is no longer sufficient.<sup>1</sup> More advanced cooling techniques entering the mainstream for personal computers include liquid cooling, ‘Peltier-effect’ based active cooling, and direct IC substrate cooling. Each of these cooling approaches presents its own set of problems that further complicate system design.

A reader who has owned a personal computer built in the late 20th century or the early 21st century understands the system constraints all too well. The noise of the computer cooling fans, the short battery life of portable computers,<sup>2</sup> and the unwanted warmth generated by the computer on hot days have been common nuisances. Worst of all, overheating of integrated circuits not only complicates the system design, it can cause the ICs to fail prematurely because of unwanted and uncontrolled diffusion of dopants, leading to shorts between source and drain regions which should remain separate.

The power-dissipation problem applies not only to computers. It is even worse for the industrial equipment used for telecommunication and networking, where ambient temperatures are consistently higher than nominal home and office room temperatures. Inefficiency of the cooling systems further compounds the problems.

The rapid depletion of our natural resources, our overloaded power-distribution grids, and the growing need for miniaturization all further underscore the importance of low-power techniques, from computers to consumer electronics such as televisions.

---

<sup>1</sup>Apple Computer’s Power Mac G5 uses liquid cooling of the microprocessor, and has four temperature zones. The temperature of each zone is controlled by individual fans and thermal sensors. This computer is surprisingly quiet despite having more than eight fans and a design resembling a wind tunnel (Dreier, 2003).

<sup>2</sup>The earlier generation of portable computers were commonly referred to as ‘laptops.’ Later the term ‘notebooks’ took over. Certainly the term ‘laptop’ was not suitable, because a number of the available models—including the Dell Inspiron 8100 used to write this dissertation—ran too hot to put on one’s lap at times!

## 2.1 History of Stanford’s Ultra-Low-Power (ULP) Technology

This dissertation is part of an ongoing effort launched at the Stanford University Electrical Engineering Department’s STAR lab, in 1990, to optimize the power dissipation of computing systems. This effort—known as Stanford’s ULP technology—was initially led by Professor Allen Peterson and James B. Burr. The idea of micropower CMOS was first described by Richard Swanson (1975) at Stanford University, where he proposed the use of ion implantation to optimize the threshold voltage of a transistor. Based on the work of Swanson, and suggestions from Professor Mark Horowitz, Burr proposed to operate circuits at optimum supply and threshold voltages using back-gate bias for threshold-voltage adjustment (Burr & Peterson, 1991; Burr, 1993; Burr, 1995a; Burr, 1995b). Supply-voltage reduction and its benefits were described at ISSCC (Burr & Shott, 1994) for  $V_{DD}=200$  mV operation of a CMOS encoder/decoder circuit. Burr also introduced a fundamental concept of balancing of active and leakage power for the most energy-efficient operation.

Burr *et al.* (1996) published a chapter on Stanford’s ULP technology in a book “Low-power HF Microelectronics: a unified approach” published by the IEE press. Professor Allen Peterson passed away in late 1994, and Professor G. Leonard Tyler took over the ULP research group and continued the research effort in collaboration with Burr, who was also a researcher at SUN Microsystems during 1995–2001. In that time period, three Ph.D. dissertations were completed in the area of low-power computation, signal processing, and circuit design including research by students Gerard Yeh, Yen-Wen Lu, and Bevan Baas. Visiting researchers Masataka Matsui from Toshiba, and Sergio Bampi from Federal University Rio Grande do Sul, Brazil, were hosted. At Stanford, Matsui designed and fabricated a 32-bit multiplier to demonstrate the effectiveness of ULP in reducing the power dissipation. Baas designed and fabricated an FFT processor, ‘SPIFFEE,’ and demonstrated the application

of ULP to signal-processing circuits. Several integrated circuits were designed to demonstrate the applications of ULP technology.

In the fall of 1996, Vjekoslav Svilan began to develop algorithms to optimize system supply and threshold voltages with back-gate biasing. He later reported important results on energy-saving measurements performed using supply-voltage scaling and threshold scaling (Svilan *et al.*, 2000). The research described in this dissertation was started in 1998 in the ULP group with Burr as the industrial mentor. The objectives of this research were to understand and optimize transistors and technology needed for ULP.

Technologies developed as part of the long research efforts carried out in the Stanford ULP research group have been made available for public use. Industry already has adopted a number of these technologies, and it is now likely that dynamic back-gate biasing and related technologies will be needed to continue increasing the circuit density and frequency of operation at the rate predicted by Moore's Law.

## 2.2 Why Worry About Power Dissipation?

The power dissipation of ICs is an important aspect of electronic system architecture. ICs typically consume significant amounts of power in any electronic system. There are other power-dissipating sources in addition to the ICs; for example, a desktop computer consumes most of its power in its display, its mechanically spinning data storage disks, and its power conversion circuits. The advent of solid-state memory and display technologies, however, may eliminate always-spinning hard disks and fluorescent illuminated displays. As a result, IC power dissipation may become the most important factor to optimize, especially for mobile and high-density computational devices.

**Table 2.1:** Energy content of batteries.

Technology	Energy Density (mW-hours/mg)
Lithium-ion chemical battery	0.3
Methanol fuel cell	3
Tritium nuclear battery	850
Polonium-210 nuclear battery	57,000

### 2.2.1 Batteries

Batteries<sup>3</sup> are essential components of mobile or remotely operated electronic devices. Longevity of the battery in a system is always highly desirable. Reduction in power dissipation in the ICs can prolong battery life. However, improvements in the battery energy density can result in even longer battery life.

New battery technologies promise to improve the energy density significantly. Especially the energy storage systems that use fuel-cell technologies, energy-scavenging techniques, and micro-electromechanical techniques using nuclear material can increase battery storage capacity markedly. Table 2.1 shows a comparison of battery options that are attractive for miniaturized and mobile systems (Lal & Blanchard, 2004). Methanol-based fuel cells offer an option for storage-capacity enhancement.

Batteries based on radioactive material can offer orders of magnitude more storage capacity, but only if safety and cost issues can be addressed. Nuclear batteries are not based on fission or fusion, but on nuclear isomers that emit high-energy particles as a source of energy. Despite these advances in battery technology, low-power ICs are still needed for circuit-density enhancement.

---

<sup>3</sup>Any electric-energy-producing apparatus is referred to here as a “battery.” Note that the benefits of reducing IC power dissipation are still relevant even given the use of advanced ‘batteries’ such as fuel cells.

### 2.2.2 Computing-circuitry density

Planar semiconductor technology integrates circuits on the surface of a wafer; planar integration includes decades-old technologies built upon the characteristics of planar ICs. Systems built with planar ICs use cooling by large heatsinks, albeit at the cost of increased system volume. However, power-efficient ICs that can enable high-computational-density designs and do not require large-surface-area heatsinks are necessary for future computing platforms.

In contrast to planar ICs, biological systems have higher computational density as compared to manmade computing machines. Biological systems are 3-D in nature, carry out multiple functions in parallel, and still operate with low-power dissipation. For example, the human brain is capable of complex functions including large data retention, sensitive sensory interfaces, complex biological controls, logic, and intelligence with an average power consumption of about 20 W.<sup>4</sup>

Three-D stacking has long been used at the printed-circuit-board level. Recently, 3-D stacking of IC packages and silicon chips has been explored. A more advanced area of research concerns 3-D stacking of the electronic devices on the silicon wafer. This solves many fundamental problems of planar technology, such as interconnection delays; but this approach presents additional challenges such as managing heat flow. Signal delays due to long-haul lateral interconnects can be reduced by vertically stacking circuits, thereby reducing the length of interconnects. In addition, vertical stacking results in increased storage density. Regarding power, existing surface cooling schemes are insufficient unless the circuits and devices are optimized for power dissipation.

---

<sup>4</sup>As an interesting comparison, based on a 2000 kCal daily diet, each human dissipates about 100 W. Each dietary kCal equals about 4000 Joules. The brain consumes about 20% of the total calories, in the form of sugars.

### 2.2.3 Environmental impact

The rapid modernization around the world, and the global thirst for information technology, are expected to increase the usage of computers worldwide. This implies global growth in energy needs due to computers, leading to depletion of fossil fuels and pollution of the environment. This problem is further compounded by inefficient and overloaded power grids.

The link between computer power efficiency and the environment becomes clear using a simple estimate. Assume that each one of 10 billion anticipated humans by the year 2050 owns a computer with a 500 W power supply, and uses this computer 20% of the time, on average. The total power needed for those computers will be about 1 TW. In 2003 roughly 210 million barrels of oil were produced per day, and given 20% conversion efficiency, this is roughly 2.9 TW,<sup>5</sup> which suggests that 1 TW of computation is very expensive. While the validity of this estimate can be challenged on many grounds, it indicates that a growing usage of computers would have a clear impact on global energy consumption.

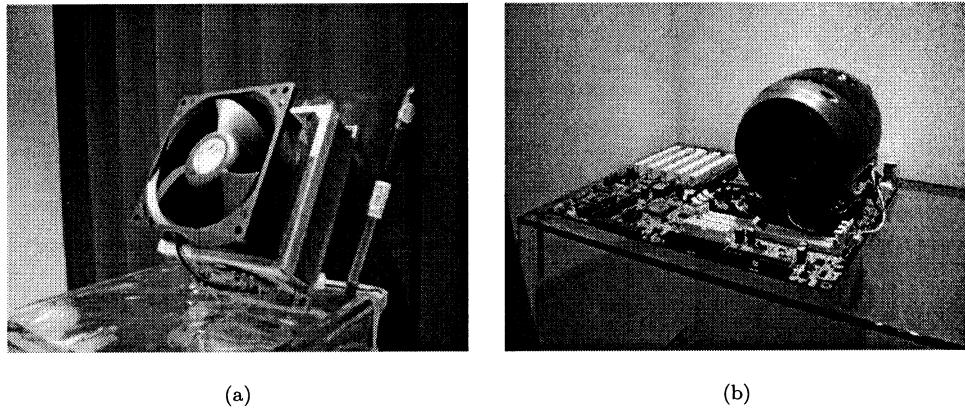
The achievement of power-efficient computing is, therefore, not only needed for the continuation of Moore's law, it is part of our collective responsibility given our environmental concerns.

### 2.2.4 System considerations

The primary considerations in a computer system design for a given performance are power dissipation, power-supply characteristics, and availability of space. High-performance systems typically employ active cooling solutions. These include forced-air cooling, liquid cooling, refrigeration, and forced-electric cooling based on the Peltier effect. Forced-air-cooled systems are easier to implement and are prevalent, but require relatively large heatsinks and have system-design limitations.

---

<sup>5</sup>Each barrel of oil contains approximately  $6 \times 10^9$  Joules.

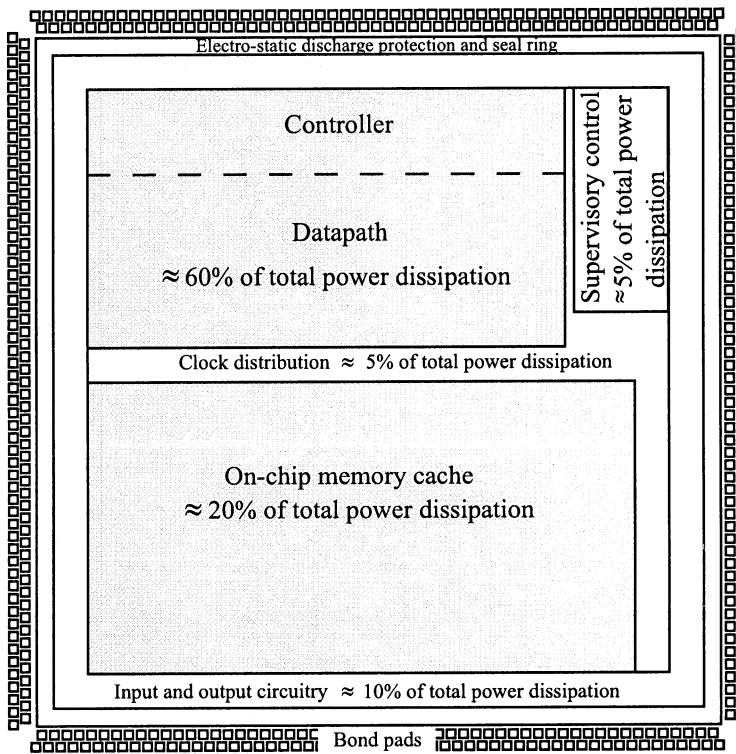


**Figure 2.1:** (a) Commercially available heatsinks for Intel microprocessor ICs. Compare with the size to the ball-point pen to the right of the fan in the picture. This heat sink is meant to cool about an  $\approx 15 \text{ mm} \times 15 \text{ mm} \times 0.5 \text{ mm}$  piece of silicon. (b) The object with the fan is another commercially available heatsink for the same chip. This heatsink appears to be huge compared to the motherboard it is mounted on. These pictures appeared in the press coverage of the Computex 2004 exhibition, May 31-June 04, 2004, Taipei, Taiwan.

Figure 2.1 shows heatsinks proposed for one of Intel's microprocessors, circa 2004. These heatsinks are space-consuming, and are incompatible with high-density or mobile designs.

Forced-liquid cooling solutions have been commercially used and researched, but pose severe mechanical and financial requirements on the system design. For example, an anti-algae agent is usually required in the liquid to prevent algae growth in the tubing. Such agents can erode tubing in the coolant distribution system, with a real risk of flooding and catastrophic loss of cooling.

Product designers want to design small, thin, robust, and rugged products. Microprocessors that run hot and require large heatsinks are not design-friendly or user-friendly. On the other hand, microprocessors that run cooler without requiring fans for heat dissipation, while still delivering high performance, are more desirable for mobile and high-density applications.



**Figure 2.2:** Example of a microprocessor. The majority of the power-dissipating circuits are digital in nature.

## 2.3 What's Inside Those Power-Hungry Integrated Circuits?

Power-hungry ICs typically include a microprocessing unit. Such microprocessors are required to deliver several billion calculations every second. The computational power of these microprocessors continues to increase over time, especially because more functionality is packed on the same dice, e.g., multiple microprocessors on a single chip. In addition to computation, microprocessors are used as embedded controllers, where typically they act as the ‘brain’ of the system and employ communication channels to every part of the system. On-chip circuitry is required to support all the above-mentioned functionality and even more.

Figure 2.2 illustrates the layout of a typical vintage 2004 processor. The core logic consists of a datapath and its controller. There is an on-chip memory cache that is used for the program and the data memory. It is, however, ‘area-expensive’ to integrate significant memory on the chip. Consequently most of the system’s memory is off-chip. High-speed I/O circuits are used to communicate with the off-chip memory as well as with the rest of the system. Clock synthesis and distribution circuitry is always needed for synchronous designs.

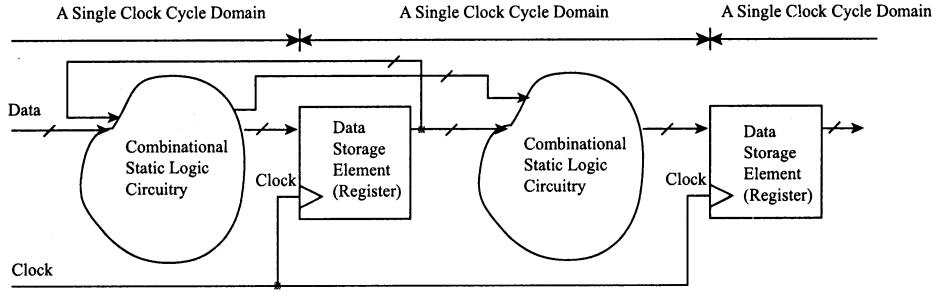
Power dissipation in ICs is concentrated mainly in digital computational, control, and memory circuits, and in analog input/output (I/O) circuits.<sup>6</sup> I/O circuits are used for inter-chip communication and inter-system communication. The signal-to-noise performance specification usually sets a lower limit on the power transmitted via a channel; although it is possible to optimize I/O circuits for power, the only effective approach to I/O power reduction is usually smart system design. Analog circuitry built in processes targeted for digital circuits is required to function across 20–30% manufacturing and operational variations. Therefore, in order to guarantee functionality across such wide variations, designers typically incorporate wide performance margins at the cost of increased power consumption. Generally, analog circuitry is a small portion of the IC, and its power consumption is a small portion of the total power.

With the exception of clock synthesis, memory, and the outward-facing-end of I/O circuits, the remainder of the functionality is implemented using digital circuits. The datapath, a digital block, is the core circuit of the microprocessor, and typically consumes a substantial fraction of the total power consumed.

Figure 2.3 shows a digital datapath, typical of a synchronous digital design style. A clock signal controls the data flow. The data move through the combinational logic to storage elements or registers. The design can include many pipelined registers.

---

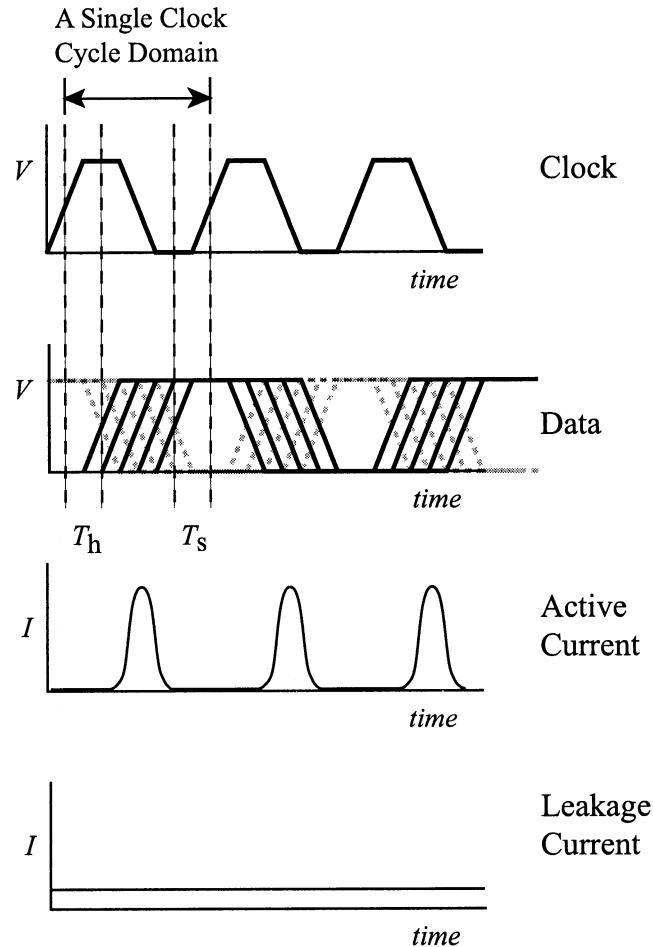
<sup>6</sup>In a sense all circuits are fundamentally analog; but circuits are called ‘digital’ when they are allowed to assume only a few (normally two) discrete states, rather than a continuous range of states.



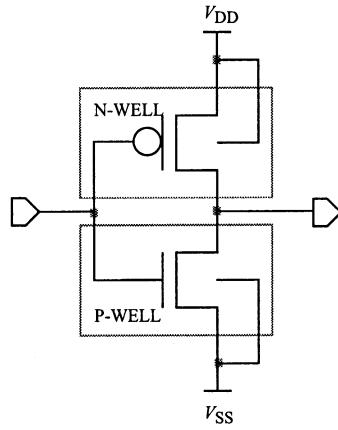
**Figure 2.3:** Typical datapath for synchronous circuit design. Data flow through a combinational circuit and then are latched into memory elements at a clock edge. Clock edges initiate a domino effect in the data switching activity.

Figure 2.4 shows typical switching activity through a datapath circuit. In this example the registers store data on every rising clock edge. The rising-edge-triggered data flow through the combinational logic before being latched into the memory elements. Such a design needs to include certain timing margins for proper registration of the data in the storage elements. Interval  $T_s$  is the minimum setup time required for the data to become stable before a rising clock edge, while interval  $T_h$  is the minimum hold time required for the data to remain stable after a rising clock edge. Note that the switching current is present only during a small period of time following every rising clock edge—this current is commonly known as the active current.

Very large scale integrated (VLSI) circuits, including transistor counts on the order of hundreds of millions, are built using complementary metal oxide semiconductor (CMOS) transistor technology. Transistors with complementary characteristics make it possible to implement circuits that have extremely low leakage current flow from power supply to ground, in order to maintain a particular stable state. The absence of any large-value continuous current path between supply and ground in CMOS—excluding leakage—makes it possible to integrate hundreds of millions of transistors onto a small silicon chip. Integration density has now reached such a high level that the power dissipated by leakage currents is a significant portion of the total power, however.



**Figure 2.4:** Switching activity through a typical datapath circuit. Rising-clock-edge triggered data-switching activity flows through the combinational logic before latching into the memory elements.  $T_s$  represents minimum setup-time required for the data to be stable before a rising clock edge.  $T_h$  represents minimum hold-time required for the data to be stable after a rising clock edge. The active current is present only during the data-switching interval.

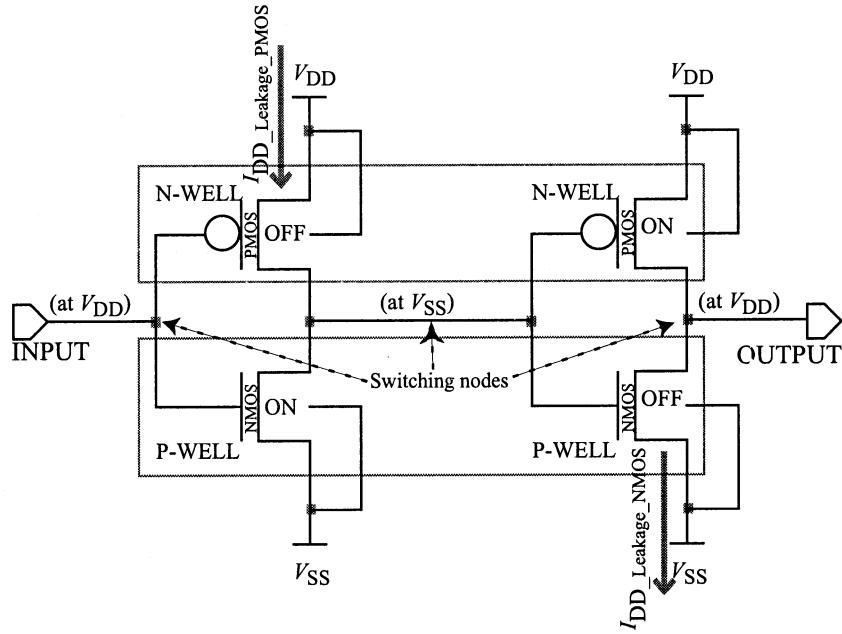


**Figure 2.5:** A static-CMOS inverter built with a PMOS transistor and an NMOS transistor. The PMOS transistor is built in an N-well and the NMOS transistor in a P-well. The P-well is biased to  $V_{SS}$  and the N-well is biased to  $V_{DD}$ . Proper well biasing is important to maintain isolation among the transistors.

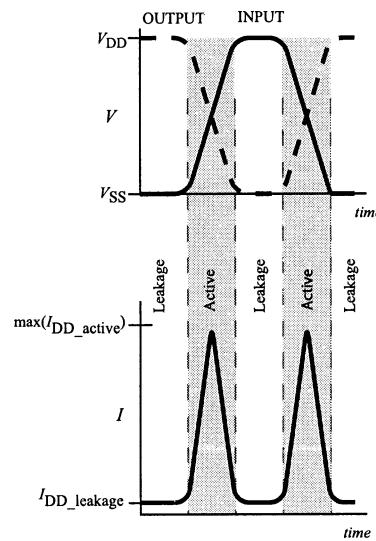
Figure 2.5 shows a CMOS inverter. A PMOS transistor is used to ‘pull up,’ and an NMOS transistor is used to ‘pull down.’ PMOS transistors are built in N-wells, and NMOS transistors in P-wells. Typically, the wells are tied to supply rails, as shown.

Figure 2.6 shows two static-CMOS inverters in series that are in a stable state. For the first inverter, the input is high ( $V_{DD}$ ), and the output is low ( $V_{SS}$ ). The NMOS transistor is on, but the PMOS transistor is off. The only current flowing through the inverter is the off-state PMOS leakage current. The input of the second inverter is low and the output is high. In this inverter, the NMOS transistor is off and the PMOS transistor is on, and therefore the current through the PMOS transistor is NMOS leakage current.

Figure 2.7 shows the currents flowing through an inverter during its switching activity. In a static state, either the PMOS transistor or the NMOS transistor is off. It is only the leakage current, of the transistor that is in the off state, that flows through the circuit in this condition. During state-switching activity, the inverter either charges or discharges its output-node capacitance. Active power is dissipated



**Figure 2.6:** A static-CMOS-inverter chain. In any static state, the only power dissipation is due to the leakage current. Note that this figure shows switching nodes to be quiescent.



**Figure 2.7:** Input and output waveforms of a CMOS inverter along with the current.

in the inverter's output resistance during a charging or discharging event, should its state change.

## 2.4 Power-Reduction Techniques for Circuits

A number of power-reduction schemes have been proposed for use at the circuit level. Among these are supply gating, clock gating, asynchronous circuits, adiabatic circuits, multi- $V_T$  static-circuit optimization, dynamic supply-voltage scaling, static supply-voltage scaling, dynamic threshold-voltage scaling, and static threshold-voltage scaling or multi- $V_T$  design. Different techniques are suitable for different applications. We list a number of parameters that are often considered in selecting an appropriate circuit technique.

### 2.4.1 Parameters to consider for power optimization

#### Optimization of CMOS

In CMOS circuits we need to optimize both leakage power and active power. The optimum power, however, may sometimes not be the minimum power in order to achieve the desired frequency of operation. Leakage power mainly comes from currents leaking between the source and the drain of a CMOS transistor in the off-state, whereas active power results from the charging and discharging of actively switching nodes. Leakage power can be reduced by increasing the threshold voltage ( $V_T$ ). Similarly, reducing the supply voltage also reduces both active power and leakage power. Increasing  $V_T$  or reducing  $V_{DD}$  slows down the circuit, however.

#### Types of data processors

Data processors can be broadly divided in two categories from the viewpoint of circuit activity: (i) fixed rate processors, and (ii) variable rate processors. Power-optimization techniques are different for these two types of processor applications.

Fixed-rate processing requires continuous computational throughput and latency. Examples of such applications include signal processors, multimedia processors, and

communication processors. The circuitry in such processors usually is designed to operate at a high, and fixed, level of activity.

Variable-rate processing is event-based or interrupt-based; the computational throughput and latency both vary with time. Most of the time such systems are idling, but for short periods of time they have high throughput. An example of such a system is the microprocessor within a notebook computer. Interestingly, the microprocessor of a notebook computer idles most of the time, under typical workloads.

### Modes of operation

At the block level, circuits are generally either idle or operational. Optimization techniques have to take into account the mode of operation. The power dissipation in the idle mode is typically due to leakage currents; therefore, power-optimization techniques suitable for leakage reduction are needed in that mode of operation. In the operational mode, a combination of architectural, algorithmic, and active-circuit power-reduction techniques are needed. It should be noted, however, that an optimum circuit and system design may result in many different modes of operation, and a suitable combination of power-optimization techniques may be warranted.

### Static vs. dynamic

The IC designer must make an important choice—whether to use a static design-optimization technique when the circuit is designed, a dynamic design optimization technique that optimizes the power at run-time, or a combination of the two. This choice is based on the circuit designer’s predictions of the usage and activity of the IC. For constant-rate microprocessors, such as signal processors, the usage and activity of the circuitry can be predicted accurately at the design stage, and so that the power can be optimized during the design process. For variable rate processors, such as notebook microprocessors, it is difficult to predict the application and usage of the circuit. For this reason, dynamic optimization control during the operation of the circuit is the best solution.

Dynamic power optimization control can be divided further into two broad categories: pre-determined open-loop, and dynamic feedback control. An open-loop

control scheme requires stable dynamic control algorithms that are well understood at design time and at production time. This control approach results in low computational overhead and is therefore easy to implement, but it cannot compensate for manufacturing variations or for operational anomalies. True closed-loop dynamic control of power optimization, on the other hand, can compensate automatically for manufacturing variations. Dynamic control entails computational overhead for the feedback control, however, and requires extensive verification for control stability across different applications.

### **Second-order effects**

Short-circuit and gate-leakage currents, and substrate leakage due to band-to-band tunneling, are secondary effects which are negligible in older IC technologies. Short-circuit currents can occur in capacitively-loaded static CMOS logic when circuit nodes are switching. Proper circuit design can reduce the short-circuit currents to negligible levels. Gate-leakage current historically has been negligible, because the transistor gate dielectric traditionally has been of high quality. With CMOS technologies shrinking below 130 nm, gate current increases due to usage of gate dielectrics thinner than 2 nm which is only a few atomic-layers thick. Use of thin gate dielectrics is needed to shrink the transistor features; films thinner than 2 nm are common. To put 2 nm in perspective, Si lattice pitch is only 0.3–0.4 nm. Electrons can directly tunnel through such thin gate dielectric films. Industrial and academic researchers are working on the use of high-k materials for the gate dielectric in order to address this problem. Gate-induced drain leakage (GIDL) and gate-side drain leakage (GDL) are the well-known forms of junction leakage. GIDL and GDL occur in highly doped two-sided pn junctions—typically, one side of a pn junction is lightly doped while the other side is highly doped. Band-to-band tunneling, trap-assisted tunneling, and thermal carrier generation are some of the underlying phenomena that produce increased reverse-biased currents in highly-doped two-sided pn-junctions. Highly-doped pn junctions are a consequence of CMOS scaling to very small sizes, and seem unavoidable when following the conventional CMOS roadmap.

### 2.4.2 Power optimization circuit techniques

Figure 2.8 shows possible optimization techniques for active and leakage power in relation to fixed-rate processors and to variable-rate processors, for operational-mode and idle-mode. These techniques include static or dynamic implementations of supply-voltage scaling, threshold-voltage scaling, supply gating, clock gating, and speed scaling. Application of these techniques can be extended to other modes of operation not covered here. Other power-optimization techniques include usage of asynchronous and adiabatic circuits.

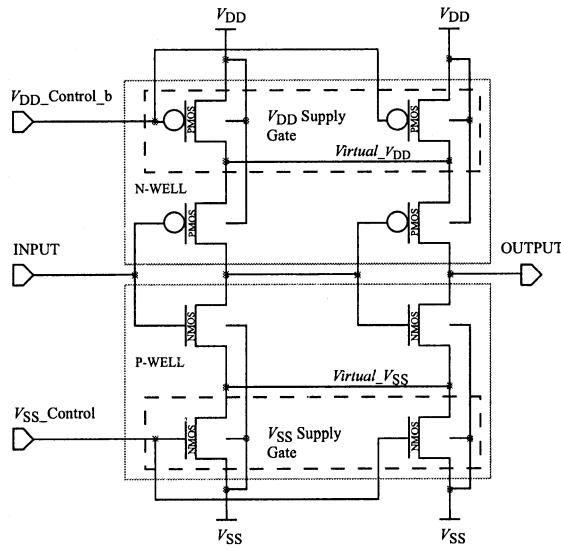
#### Supply gating

Supply gating is a circuit technique to reduce leakage power, where the reduction is achieved by inserting additional transistors in series with the power supply as well as with ground. These additional transistors increase the effective channel length of the transistor during the off state, thereby reducing the leakage current. This reduction in leakage current comes at the cost of frequency of operation, however; the voltage drop across the control transistors results in a reduction in the voltage applied to the gates of the active transistors, that in turn results in less current through the active transistors. Therefore, it takes longer to charge and discharge the nodes, reducing the frequency of operation.

Figure 2.9 shows an example of two chained inverters with parallel-strapped supply-gating transistors, located at the top and the bottom of the figure, controlling the voltage on virtual power rails  $\text{Virtual-}V_{DD}$  and  $\text{Virtual-}V_{SS}$ . In idle mode, the control transistors are off and the virtual rails float. Although the leakage current is reduced as an effect of additional supply-gate transistor, the output of the circuit cannot be predicted due to the presence of the floating virtual power-supply rails. The uncontrolled voltage of the virtual power rails during the idle mode also results in the loss of any data in a storage element that is powered from these rails. This problem is addressed by using extra storage elements that are connected directly to the power supply. Addition of extra storage elements and control transistors requires additional area on the IC, and introduces a “powerup” latency.

Static Design	1- Supply-voltage scaling (static)	1- Threshold-voltage scaling (static) 2-Supply gating	
Dynamic Design			
	Active power	Leakage power	
(a) Operational-mode optimization techniques for constant rate processors e.g., signal processors.			
Static Design	1- Clock gating	1- Supply-voltage scaling (static) 2- Supply gating	
Dynamic Design		1- Threshold-voltage scaling (dynamic) 2- Supply-voltage scaling (dynamic)	
	Active power	Leakage power	
(b) Operational-mode optimization techniques for variable rate processors e.g., microprocessors.			
Static Design	1- Clock gating	1- Supply-voltage scaling (static) 2- Supply gating	
Dynamic Design		1- Threshold-voltage scaling (dynamic) 2- Supply-voltage scaling (dynamic)	
	Active power	Leakage power	
(c) Idle-mode optimization techniques for constant rate processors e.g., signal processors.			
(d) Idle-mode optimization techniques for variable rate processors e.g., microprocessors.			

**Figure 2.8:** Application space for power-reduction circuit techniques.



**Figure 2.9:** Supply gating. Series control transistors gate the supply voltage and the ground connections. The control transistors reduce the source-to-drain leakage significantly by breaking the current loop.

### Clock gating

Power consumption in synchronous circuit design can suffer from unnecessary circuit activity such as continuous clock distribution. The clock distribution to all the storage elements results in node toggling, even though not all the circuitry is being used all the time. Output may not be needed in all portions of the circuitry; therefore, some portions of the circuitry may not need to receive the clock under certain operating conditions. The clock input to unneeded circuits can be suspended or shut down, avoiding unnecessary charging and discharging of nodes and improving efficiency. Clock-gating is a commonly used technique to use this objective; it is now a standard practice to incorporate this feature in digital circuit design.

### Asynchronous design

Synchronous design, the dominant digital circuit design technique, is based not only on quantized voltage or current levels but also on quantized time steps. An entirely different approach for digital circuits, commonly known as asynchronous design, does not require quantized time steps. Asynchronous circuit design has been around for a

long time, but has not received much attention due to the fact that it is difficult to implement and more error-prone as compared to synchronous design.

Unlike synchronous circuits, asynchronous circuits do not depend on a global clock. The absence of the unnecessary clocking and circuit activity often results in relatively lower power requirements. There may also be speed benefits, because component idle time near the end of each clock cycle is sometimes reduced. Asynchronous systems are modular, and deliver ‘average’ performance as opposed to the worst-case performance for synchronous designs (Myers, 2001). Asynchronous design has not been widely used in industry; it is, however, an interesting topic, and more research is needed in this area in order to demonstrate practical implementation and utility.

### **Adiabatic circuits**

It has been shown that, if a computation is logically reversible and can be implemented in a physically reversible technology, then computing engines can be designed so that they do not require energy dissipation (Merkle, 1993). Circuit families that implement this reversibility concept are referred to as *adiabatic circuits*. In simple words, the energy used for one state transition is recovered during the next state transition. Adiabatic circuit design styles—and there are many—operate at slow speeds, which makes them ideal for applications, such as cardiac pacemakers, where power dissipation must be in the  $\mu\text{W}$  range.

### **Dynamic speed scaling**

Active power dissipation, which results from switching activity, is proportional to frequency of operation. By reducing the speed of operation, the level of required input power can be reduced. In CMOS circuits, dynamic speed scaling results in power savings when used in conjunction with dynamic supply-voltage scaling. Reduction of the supply voltage usually results in reduction of the maximum speed of operation for a CMOS circuit. At the same time active power is reduced, not only from frequency reduction, but also because the lower supply voltage yields a reduction in leakage.

### Supply-voltage scaling

Supply voltage ( $V_{DD}$ ) is an important factor in power dissipation. Increasing supply voltage increases the active power dissipation as well as the leakage currents. For this reason it is important to operate a chip at an optimum supply voltage for a given speed performance. The implementation of this concept can be either dynamic or static.

In a static design approach, i.e., when power optimization occurs only at the design time, multiple power-supply domains can be made available on the IC. ‘Slower’ circuitry is powered at a lower supply voltage, whereas ‘faster’ circuitry is powered at a higher supply voltage. Such a scheme works well for fixed-rate processing designs in the active mode as well as in the idle mode.

Dynamic control of supply voltages can be used to effectively optimize power dissipation at runtime for a variable rate application, such as a microprocessor for a notebook computer. When a system is in the idle mode, there is no need to have the supply voltage as high as it needs to be for its highest-speed operation. In an ideal implementation, the supply voltage should be adjusted to optimize power dissipation for each frequency of operation.

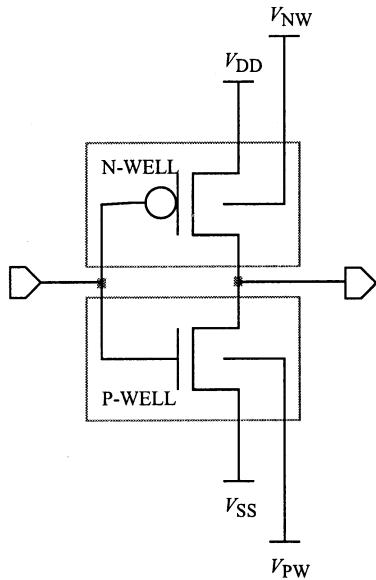
Burr (1993) demonstrated operation of CMOS circuits at  $V_{DD}=200$  mV, pointing out the possibility of power savings for very low voltages. Svilan (2000) demonstrated operation of a multiplier circuit at even lower voltages. Dynamic voltage-scaling technology was first commercially demonstrated by Transmeta Corporation (Ditzel, 2000). Every major microprocessor design now uses voltage scaling in some form.

### Threshold-voltage scaling

Modern IC technologies offer a variety of transistors with a choice among multiple threshold voltages. Transistors with higher threshold voltage ( $V_T$ ) leak less as compared to transistors with lower  $V_T$ ,<sup>7</sup> but lower  $V_T$  transistors run faster than higher  $V_T$  transistors. If the circuit application is well known in advance, for instance

---

<sup>7</sup>Leakage currents related to threshold voltages are referred to as subthreshold leakage currents and are described in Chapter 3.



**Figure 2.10:** A CMOS inverter with separated back-gate terminals. In the conventional CMOS technology,  $V_{NW}$  controls the back-gate bias on n-wells and  $V_{PW}$  controls the back-gate bias on p-well or p-substrate.

for network processing, it makes sense to optimize the circuits for power dissipation at design time. In this approach, the transistors that are not critical for speed, and for which leakage is the primary cause of power dissipation, should be manufactured with higher  $V_T$ . Transistors for which operating speed is critical can be manufactured with lower  $V_T$ . This design methodology requires design tools that can identify slow and fast datapaths and use transistors with appropriate  $V_T$  for each path. Multi- $V_T$  static design optimization is expensive in terms of tools and design effort, and can result in a number of speed-limiting paths. This approach is best suited for systems which perform predictable tasks. Multi- $V_T$  may not be a suitable option for high-performance microprocessors, because of the diversity of applications in which they are employed.

Threshold tuning with back-gate bias<sup>8</sup> does not require any significant change in circuit design style. System power dissipation is optimized by balancing active and

---

<sup>8</sup>The terms “body-bias,” “substrate-bias,” “back-bias,” “back-gate-bias,” and “bulk-bias” all refer to the same concept and the same technique.

leakage power, which is achieved by optimizing supply and threshold voltages. The threshold-voltage adjustment is achieved by using back-gate bias.

*Now, having briefly toured the status of low-power design practice, we next zero in on the device requirements for the implementation of ULP circuit techniques.*

## Chapter 3

# Device Requirements for Ultra-Low-Power Technology

In this chapter, we describe requirements for ULP device technology. First, we discuss the sources of power dissipation in CMOS circuits. We then present Burr's Law of ULP, and its implications with regard to using power-supply and threshold-voltage adjustment. We conclude the chapter by outlining the desired characteristics of the devices under investigation.

Ultra-low-power (ULP) technology encompasses circuit and device techniques intended to achieve the most energy-efficient operation possible. One important condition for the most energy-efficient operation of digital CMOS circuits is that dynamic power dissipation due to circuit switching is essentially equal to the static power dissipation due to leakage. ULP technique dictates the adjustment of the power-supply voltage and the threshold voltage of MOS transistors to achieve this balance. Therefore, the implementation of ULP device technology requires a mechanism for dynamic adjustability of this threshold voltage.

### 3.1 Power Dissipation in CMOS

Power dissipation in CMOS circuits is divided mainly into three categories: active, short-circuit, and leakage.

### 3.1.1 Active power dissipation

Active power is that dissipated due to the transient charging and discharging of parasitic load capacitances when the circuit switches its state, and is proportional to circuit activity, switching frequency, load capacitance, and the square of supply voltage. The active power is given by

$$P_{\text{active}} = aC_{\text{load}}V_{\text{DD}}^2f \quad (3.1)$$

where  $a$  is the circuit activity factor,  $C_{\text{load}}$  is the equivalent load capacitor,  $V_{\text{DD}}$  is the power-supply voltage, and  $f$  is the frequency of operation. This formula, however, gives no direct information about the device design.

Device engineers attempt to maximize the current through the transistor channel ( $I_{\text{DS}}$ ) under conditions corresponding to the maximum gate voltage; the maximum gate voltage is the same as  $V_{\text{DD}}$  for static CMOS logic. Higher values of  $I_{\text{DS}}$  result in quicker charging and discharging of the nodes, and therefore higher frequency of operation. A model of this current is described by (Burr *et al.*, 1996)

$$I_{\text{DS}|V_{\text{DS}} \leq V_{\text{Dsat}}} = \mu_{\text{eff}} C_{\text{ox}} \frac{W_{\text{eff}}}{L_{\text{eff}}} (V_{\text{GS}} - V_{\text{T}} - \frac{V_{\text{DS}}}{2}) V_{\text{DS}} \frac{1}{1 + \frac{V_{\text{DS}}}{E_{\text{sat}} L_{\text{eff}}}} \quad (3.2)$$

Here  $\mu_{\text{eff}} C_{\text{ox}} \frac{W_{\text{eff}}}{L_{\text{eff}}}$  is a device-technology constant,  $V_{\text{GS}}$  is the voltage across gate and source terminals, and  $V_{\text{DS}}$  is the voltage across drain and source terminals of the transistor. For details about the parameters in this equation, please refer to Appendix A.

Note that in addition to the technology constant,  $I_{\text{DS}}$  strongly depends on  $V_{\text{GS}} - V_{\text{T}} - \frac{V_{\text{DS}}}{2}$ , commonly known as overdrive voltage, and  $V_{\text{DS}}$ . The maximum value of the overdrive voltage can be  $V_{\text{DD}} - V_{\text{T}}$  and the maximum value of  $V_{\text{DS}}$  can be  $V_{\text{DD}}$ . Attempting to achieve higher  $I_{\text{DS}}$  by increasing  $V_{\text{DD}}$  can achieve higher frequency of operation, but at the cost of increasing power dissipation.

The power-to-frequency ratio can be optimized by reducing  $V_{DD}$  while maintaining  $V_{DD} - V_T$ , and increasing the device-technology constant so that the maximum value of  $I_{DS}$  either increases or stays unaffected. This approach typically results in increased maximum frequency of operation for a constant power dissipation. It may seem counterintuitive, but a transistor technology that results in higher values of  $I_{DS}$  for a lower  $V_{DD}$  may be more suitable for low-power operation. One way to understand this is to think in terms of transistor channel resistance: higher current for lower overdrive voltage implies lower transistor channel resistance, and therefore less power dissipation in the channel due to high resistance. It has been, therefore, the trend to shrink the transistor channel length and lower the supply voltages accordingly. We also present an edge-defined nanolithography technique in Chapter 5 to make short-channel devices, which when combined with mobility improvement techniques has the potential for drive-current enhancement.

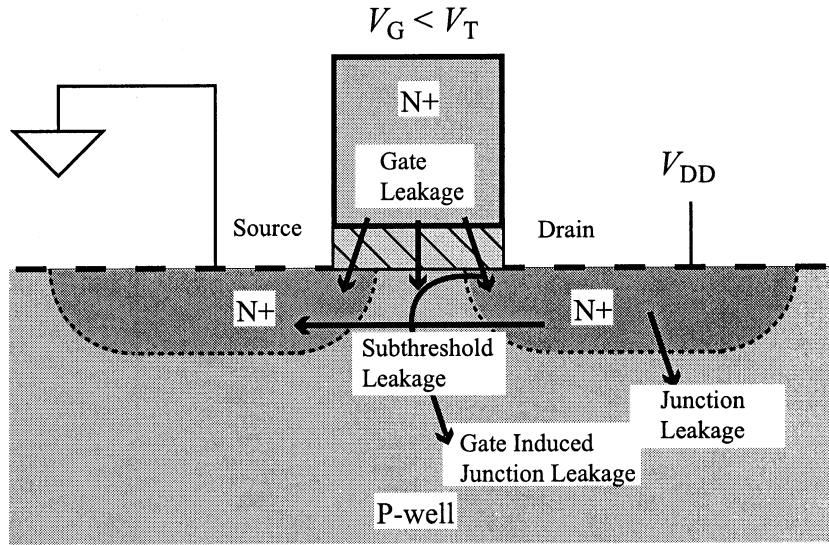
### 3.1.2 Short-circuit power dissipation

During circuit switching, a momentary short circuit can occur between power supply and ground if the pull-up and pull-down transistors are turned on simultaneously. With proper, common-practice circuit design, short-circuit power dissipation is negligible. We, therefore, ignore the short-circuit power—it is an issue to be addressed at the circuit-design level.

### 3.1.3 Leakage power dissipation

Leakage current flows between different transistor terminals due to several underlying phenomena. Figure 3.1 illustrates the leakage current paths in a turned-off transistor. Leakage current has three components: subthreshold current, gate tunneling current, and junction leakage current due to band-to-band tunneling as well as gate-induced drain leakage.

Subthreshold current is the dominant leakage component. Figure 3.2 shows typical turn-off characteristics for a MOS transistor. For  $V_G < V_T$ , the drain current decreases



**Figure 3.1:** Cross-section of a turned-off NMOS transistor. Current leaks through the channel, the gate oxide, and the reverse biased junctions. Junction leakage is significant near the channel because of the heavily doped two-sided p+/n+ junctions. It is negligible away from the channel because the junctions only have one heavily doped side away.

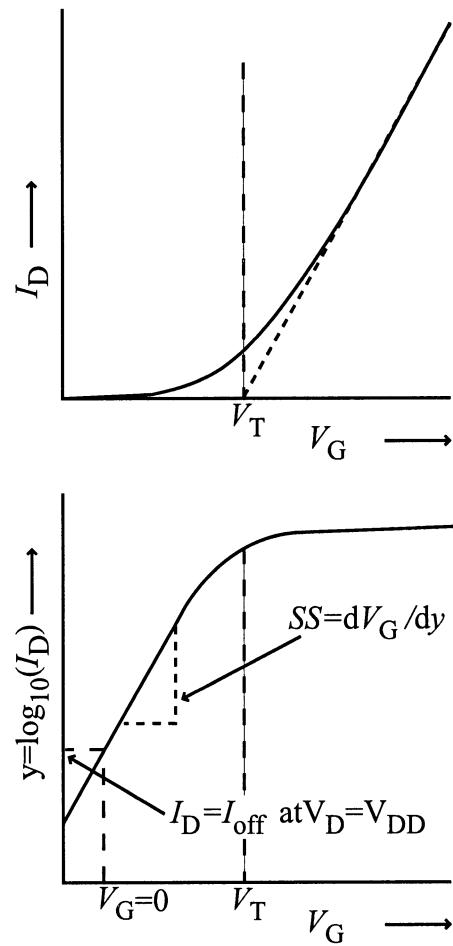
by 10× for every 70–100 mV reduction in  $V_G$ —a rate commonly referred to as the subthreshold slope ( $SS$ ). For  $V_G = 0$ ,  $I_D = I_{\text{off}} > 0$ , and the objective of the device designer is to reduce the value of  $I_{\text{off}}$ . By looking at Fig. 3.2, it is clear that a steeper slope in the turn-off region results in lower  $I_{\text{off}}$ .

Subthreshold slope depends on the transistor structure. For a traditional structure,  $SS$  is given by

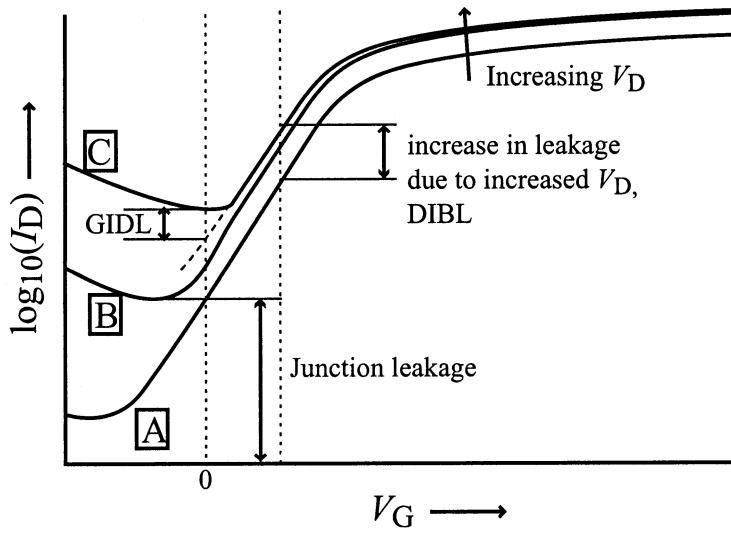
$$SS = \frac{kT}{q} \ln(10) \left( 1 + \frac{C_{\text{dep}}}{C_{\text{ox}}} \right), \quad (3.3)$$

where  $C_{\text{dep}}$  is the capacitance in the depletion region of the channel, and  $C_{\text{ox}}$  is the capacitance of the gate oxide. Note that this model predicts  $SS \geq 60$  mV/decade in a limiting case assuming room temperature.

The gate terminal of an MOS device has traditionally been predominantly capacitive in nature. With shrinking gate-dielectric film thicknesses, however, the leakage DC currents through the gate dielectric are not negligible anymore.



**Figure 3.2:** Typical turn-off characteristics of a MOS transistor. It is more common to plot  $I_D$  on the log scale in order to observe the turn-off behavior.

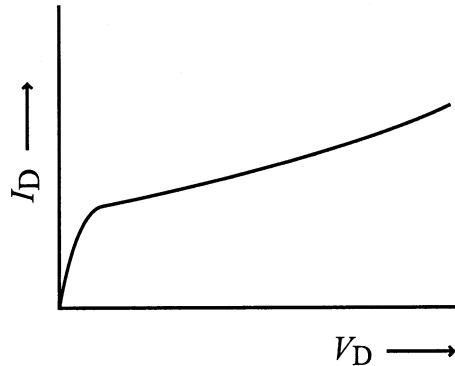


**Figure 3.3:** Typical turn-off behavior of a MOS transistor for three distinct values of drain voltages.

Figure 3.3 describes all the leakage components of the drain currents. Curve A is for  $V_D \leq 0.1$  V; at this low value of  $V_D$ , the drain voltage does not contribute to the second-order effects and the only leakage current is of the subthreshold type. Curve B is for a higher value of  $V_D$ ; the minimum leakage current, towards the left side of the curve, is higher than the minimum leakage current for curve A. This increase in the minimum leakage current is due to the drain-junction leakage, which has higher reverse voltage across it as compared to that of curve A. Drain-junction leakage current increases with increasing drain voltages. When  $V_D = V_{DD}$  and the gate voltage is decreased, a localized depletion zone occurs near the Si surface of the drain junction. This creates a localized highly doped junction with reverse-bias voltage, resulting in increased leakage current—known as gate-induced drain leakage (GIDL). The underlying phenomena for junction leakage and GIDL are the same, therefore we include them under the same category.

### 3.1.4 Drain-induced barrier lowering

High drain voltages change the potential of the transistor channel on the drain side in a manner similar to the effect of the gate voltage, and also lower the threshold



**Figure 3.4:** Typical  $I_D$ - $V_D$  characteristics of a MOS transistor. Note that the current does not saturate for large values of  $V_D$  due to DIBL.  $V_S = 0$ .

voltage. This phenomenon, therefore, increases the drain current, and is commonly known as drain-induced barrier lowering (DIBL). Figure 3.3 shows increase in the drain current due to increased  $V_D$ .

Figure 3.4 shows typical  $I_D$ - $V_D$  curve for MOS transistors. Note that  $I_D$  shows dependence on higher values of  $V_D$ ; the increase is due to the barrier lowering due to drain-voltage coupling to the channel.

## 3.2 ULP Techniques

ULP technology derives from balancing the active and the leakage power (Burr *et al.*, 1996). The active power is adjusted by dynamically adjusting  $V_{DD}$ , while the leakage power is adjusted by dynamically adjusting  $V_T$  (Svilan *et al.*, 2000). Back-gate biasing is used to adjust  $V_T$  (Burr, 1995b). The power control is driven based on Burr's Law of ULP,

$$\frac{I_{on}}{I_{off}} = \frac{l_d}{a} \quad (3.4)$$

where,

$I_{on}$  = Drive current of a transistor turned-on,

$I_{off}$  = Leakage current of a transistor turned-off,

$l_d$  = Logic depth of the circuit, measuring the maximum number of gates in series,

$a$  = Activity of the circuit measured as the percentage of time a circuit is active.

This law states that the most energy-efficient operation of a CMOS circuit with logic depth of  $l_d$  and activity  $a$  is achieved when the  $I_{\text{on}}$  to  $I_{\text{off}}$  ratio is adjusted to be equal to  $\frac{l_d}{a}$  or vice versa. The  $I_{\text{on}}$  to  $I_{\text{off}}$  ratio can be adjusted in many different ways; during fabrication by means of static  $V_T$  adjustment using ion implantation, or dynamically at runtime by applying back-gate bias. Although supply voltage also affects this ratio, the effect is to a smaller degree. Burr's Law introduces a design target for balancing the active power and the dynamic power in order to minimize the energy of computation in CMOS circuits.

## 3.3 Characteristics of ULP Devices

### 3.3.1 Threshold-voltage tunability

Threshold voltage ( $V_T$ ) is a commonly used, extracted parameter which does not have a standard physical definition. Threshold voltage, however, is a very powerful concept that describes the gate-voltage value that acts as a threshold between the channel turn-on and turn-off. Increasing the voltage on the gate electrode of a transistor depletes the p-doped channel of its free majority-carrier holes. When gate voltage exceeds a threshold, a sheet of minority-carrier free electrons forms on the surface of the Si channel and gate-oxide interface. This sheet of charge, known as the inversion layer, results in charge conduction and thus the turn-on of the transistor. Given that there are many ways to define  $V_T$ , its absolute value is not of great significance; however, ( $\Delta V_T$ ), the change in  $V_T$  between two operating configurations, is extremely interesting and is of primary concern for ULP applications.

Adjustment of the threshold voltage is achieved via three techniques: (1) the dopant and charge trap concentration in the channel, (2) the work-function of gate material, and (3) the back-gate bias. For design-time  $V_T$  adjustment, dopant and fixed-charge concentration and work-function engineering are used. For dynamic  $V_T$  control at run-time, back-gate bias is the only option. The change in  $V_T$  for applied

back-gate bias is given by

$$\Delta V_T = \gamma(\sqrt{2|\phi_p| + |V_S - V_B|} - \sqrt{2|\phi_p|}) \quad (3.5)$$

where

$$\gamma = \frac{\sqrt{2\epsilon_s q N_a}}{C_{ox}}. \quad (3.6)$$

$\phi_p$  represents the bulk potential with respect to the Fermi potential, and depends on the bulk doping  $N_a$ .

Optimizing ULP devices requires good threshold-voltage adjustability with applied back-gate bias. Device structures that lack a mechanism for such tunability are not suitable for ULP applications; these include SOI transistors, FinFETs, surround-gate FETs, and pillar transistors. It has been noted, however, that traditional bulk transistors do not show good  $V_T$  tunability (Keshavarzi *et al.*, 1999) at CMOS technology scales below 130-nm nominal-feature size. Practical implementation of ULP technology requires new device structures that show  $V_T$  tunability for nanometer-scale devices.

### 3.3.2 Turn-off behavior

ULP technology requires that device structures attain low subthreshold slope. Low values of subthreshold slope result in higher values of  $I_{on}/I_{off}$ , which results in a wider device-adjustability range. At the same time, elimination of DIBL and depletion punch-through beneath the channel are also desirable. Ultra-thin-body (UTB) device structures address these issues by employing very thin channel films. Equation 3.3 shows that lowering the value of  $C_{dep}$  results in lower SS. Thin channel films in UTB result in low values of  $C_{dep}$ , and have the potential for lowering SS. Similarly, UTB has the potential to achieve proportional geometrical scaling in all three dimensions, which results in the elimination of short-channel effects, such as DIBL and punch-through, that result from disproportionate device shrinking.

In a traditional sense, UTB structures do not have a back-gate. However, for ULP applications, a UTB structure with a back-gate is a choice of device structure that addresses the issues related to turn-off behavior, as well as short-channel effects.

### 3.3.3 3-D device stacking

Vertical stacking of device layers can result in very promising low-power solutions when combined with ULP. Stacking of integrated-circuit device layers can greatly increase density and reduce cost, but suffers from reduced heat-removal capacity. The most important technical advantage offered by 3-D stacking is the reduction of interconnect delay (Saraswat *et al.*, 2000; Souri *et al.*, 2000). Manufacturing technology exists to make 3-D ICs commercially available for relatively slow-speed, high-density applications (Lee, 2002); however, fabrication of higher speed 3-D ICs still requires technical solutions for heat dissipation and formation of single-crystal silicon for transistor channels within a 3-D structure. This makes ultra-low-power circuit techniques an ideal technology for high-speed 3-D device stacking. Therefore, our development of the technology is targeted for compatibility with 3-D processing techniques.

*Because ULP circuit techniques are most important at extremely small-geometric feature sizes, we now explore what happens when surface irregularities loom sufficiently large in relation to feature sizes that they can affect circuit behavior, and can no longer be considered negligible. Fortunately, work-arounds are available to mitigate these effects, even though they can never be eliminated entirely.*

## Chapter 4

# Thin-Film Surfaces for ULP Device Technology

We begin this chapter with an overview of the film-deposition techniques of interest, followed by a description of the experimental details and the results. We then present surface scans of the SiO<sub>2</sub> substrate surfaces, and describe our model for film growth. We also investigate the relationship of roughness to deposition progress and processing parameters, substrate film roughness, predeposition cleaning, and an ‘etch-back’ smoothing process.

Most of the layers within a semiconductor device are, at some point during the IC fabrication process, surfaces. With layer thicknesses having been reduced to a few atoms, surface defects can seriously alter the properties of individual transistors, generally not in desirable ways. We therefore now explore technical approaches to measuring, and then to reducing, surface roughness.

Film deposition is a basic semiconductor manufacturing process. Ultra-low-power (ULP) device technology development requires manufacturing and optimizing devices with feature sizes in the nanometer-scale range. For a ULP device technology—and for continued reduction of device gate lengths ( $L_g$ ) in general—nanometer-scale roughness of deposited silicon (Si) and silicon dioxide (SiO<sub>2</sub>) films becomes significant. Si films with thicknesses of a few nanometers are useful for semiconductor fabrication

of devices for 3-D stacking. Si films just a few nanometers thick are also commonly employed for their optical properties.

Controlled lateral crystallization can be a very useful technique applicable to high-performance-transistor stacking. This process requires amorphous Si (*a*-Si) deposition for the transistor channel prior to the crystallization step (Subramanian, 1998; Joshi & Saraswat, 2003). In addition, *a*-Si can be deposited at temperatures as low as 450 °C, thereby making *a*-Si a very suitable material for device stacking.

Device-stacking processes require that the absolute maximum temperature should not exceed ≈600 °C prior to final dopant activation and crystallization annealing—assuming that the metal used for interconnects can even sustain 600 °C. Typically 450 °C is the maximum temperature allowed after aluminum metal processing. Excessive heating results in uncontrolled dopant migration in the underlying stacked device layers, making early exposure of the wafers to high temperatures detrimental to device performance or even causing outright failure. For the same reason, it is not suitable to deposit stacked transistor channels in polycrystalline Si (poly-Si), since doing so would require high-temperature processing at each new level, with deleterious effects on the dopant profile in previously worked levels.

Optimal ultra-thin-body (UTB) devices with  $L_g < 50$  nm require <10 nm Si films for device channels. Typically, UTB channels are made from single-crystal silicon-on-insulator (SOI) films. For 3-D stacking, however, low-pressure, chemically-vapor-deposited (LPCVD) *a*-Si typically is used for the channels. Successful stacking of UTB transistors requires deposition of a thin, smooth *a*-Si film on an insulator that can be used to form the transistor channel. In such applications it is important that the *a*-Si films be as smooth as possible, as the presence of roughness decreases the carrier mobility in the channel, degrading device performance (Uchida *et al.*, 2003; Uchida *et al.*, 2004). Surface-roughness characterization is, therefore, important for the formation of channels in thin-film transistors (TFTs).

Deposited films also are used to form spacer masks for edge-defined lithography. Roughness in the spacer films and on the sidewalls contributes to variations in the critical dimension (CD) of edge-defined features (Nasrullah *et al.*, 2003).

In a conventional sense, *a*-Si has always been thought of as a very smooth film. We have found, however, that this is not necessarily the case in nanometer-scale processes. For this reason we explore the impact of different parameters on the surface roughness at that scale. To this end we use an atomic force microscope (AFM) to investigate the surface roughness of Si and SiO<sub>2</sub> films having thicknesses within the range of interest for fabrication of UTB channel films. The primary objective of this surface study is to understand the nanometer-scale surface roughness on lateral scales comparable to transistor dimensions.

## 4.1 Overview of Thin-Film Deposition Processes

Methods of thin-film deposition are essential to semiconductor manufacturing processes. Such processes typically are divided into two broad categories: chemical vapor deposition (CVD), and physical vapor deposition (PVD). In a typical deposition, a clean substrate is subjected to a controlled atmosphere containing growth-promoting species. A combination of pressure, temperature, gas composition, and ambient ingredients determines the growth properties. The resulting film growth can be either the result of a chemical reaction or of a physical phenomenon.

For our applications, we are concerned mainly with the growth of *a*-Si on SiO<sub>2</sub>, as SiO<sub>2</sub> is often used as the underlying layer or substrate for *a*-Si deposition in 3-D stacking. It is important, therefore, to understand the deposition and surface roughness of the SiO<sub>2</sub> substrate as well. SiO<sub>2</sub> can be grown either by thermal oxidation processing or by LPCVD, while deposition of *a*-Si can be carried out either by LPCVD or by evaporation of Si. LPCVD of *a*-Si offers a number of advantages; it is commonly available, it is well understood, and it is a clean deposition technique. Deposition

using evaporation of the source material is a type of PVD, and is generally associated with micro-electro-mechanical systems and optical systems.

#### 4.1.1 Thermal growth of SiO<sub>2</sub>

Thermal growth of SiO<sub>2</sub> is implemented by heating the substrate at a high temperature in a furnace filled with a combination of clean oxygen gas and water vapor, which are the oxidizing species. Clean water vapor is created with a hydrogen and oxygen torch located inside the furnace. A heated surface of Si oxidizes to form a high-quality SiO<sub>2</sub> dielectric film. SiO<sub>2</sub> films grown in water vapor and oxygen are referred to as “wet-oxide.”

SiO<sub>2</sub> films grown in pure oxygen are referred to as dry-oxide. Dry-oxide is a very high-quality dielectric with few defects, and has traditionally been used as the dielectric for transistor gate electrodes. Dry-oxide grows at a slower rate than does wet-oxide, however.

Thermal oxidation of a single-crystal substrate results in a surface smoothness comparable to the smoothness of the underlying single-crystal substrate. Such thermally grown oxide films are not suitable for stacked devices, however, due to the required long exposure to high temperatures, usually above 700 °C.

It is also possible to oxidize surfaces of LPCVD Si films for transistor fabrication (Miyasaka *et al.*, 1997). Again, these processes are infeasible due to high processing temperatures.

#### 4.1.2 Low-pressure chemical vapor deposition

Chemical-reaction-based film growth processes are often used for deposition of Si, SiO<sub>2</sub>, and silicon nitride (Si<sub>3</sub>N<sub>4</sub>) films. With these processes it is possible to produce ‘device-quality’ films at low temperatures. Energy for the chemical reactions is supplied by the LPCVD reactor heating system, while reaction chemicals are injected into the LPCVD reactor at a low pressure which is maintained by a vacuum pump.

Silane ( $\text{SiH}_4$ ), disilane ( $\text{Si}_2\text{H}_6$ ), and dichlorosilane ( $\text{SiCl}_2\text{H}_2$ ) gases are typical sources of Si in such processes.

The deposition of *a*-Si occurs through a series of reactions.  $\text{SiH}_4$  partially dissociates to give  $\text{SiH}_2$  and hydrogen ( $\text{H}_2$ ). During adsorption, the Si compound molecules land on the substrate surface and diffuse to low-energy sites. The adsorbed species then decompose into Si and hydrogen. Finally the released hydrogen desorbs into the atmosphere, leaving behind deposited Si.

On a clean  $\text{SiO}_2$  substrate, nucleation begins at sites where clusters of diffusing species form, as described by Kamins (1998). The nuclei then grow at the cluster sites and coalesce into a film. Typically LPCVD Si films grown below the transition temperature of  $\approx 575^\circ\text{C}$  are *a*-Si, while Si films grown above the transition temperature are poly-Si. The exact transition temperature between *a*-Si and poly-Si depends on a number of deposition parameters (Kamins, 1980).

#### 4.1.3 Deposition by ‘evaporated’ Si

Deposition of evaporated Si on a target wafer is a type of physical vapor deposition. Paradoxically, films deposited using this method are commonly referred to as ‘evaporated’ films. This process is initiated when a source of Si in the form of a pellet is heated by an electron beam in a vacuum. As a result, Si atoms evaporate into the thin atmosphere, and then deposit by condensation on the target wafers, which are independently maintained at room temperature. The deposition is purely physical as no chemical reactions are involved. Due to the low substrate temperature, when atoms or molecules land on the surface they do not have sufficient energy to diffuse over long distances. This circumstance results in uniform deposition over the substrate.

## 4.2 Experimental Details and Results

We are now going to study LPCVD Si films on SiO<sub>2</sub>. Underlying SiO<sub>2</sub> films explored here are thermally grown as well as LPCVD. In addition we also provide results for ‘evaporated’ Si film surfaces.

### 4.2.1 Sample preparation, scanning, and characterization

Our work is performed at the Stanford Nanofabrication Facility. LPCVD is carried out in a Tylan LPCVD reactor tube, while thermal oxidation is carried out in a Tylan oxidation furnace tube. Prime quality, polished Si wafers are used as ‘handle-wafers’ for this experiment.

A Dimension 3000 AFM operating in its ‘tapping’ mode is used to scan the sample surfaces. The AFM tips used for these scans are silicon probes with specified tip radii of less than 10 nm. We have, however, observed that AFM tips wear out rapidly, and that each batch of fresh AFM tips includes a distribution of sharp and blunt tips. Therefore, we select AFM probe tips with special care, and attempt to use the same AFM tips for all scans where a consistent comparison is required.

Si samples for AFM scanning are prepared by heating at 150 °C in an oven for 30 minutes. AFM scans of Si surfaces, performed without preheating in the oven, show a fluidic film which masks the real surface topography. It may be that water adsorbed on the film surface masks the topography.

A combination of parameters is required to accurately describe the surface. For example, peak-to-peak surface roughness ( $\Delta h_{p-p}$ ) defines the extreme height variation of surface topography. Although this parameter does not characterize the surface transition profile between the maxima and the minima, it is the best single description of the surface. As we show in subsequent sections, AFM surface scans of *a*-Si films at nanometer-scale magnification reveal surface textures dominated by the presence of scattered summits. ‘Smoother’ surfaces have relatively broad areas at the summit bases, whereas ‘rougher’ surfaces have sharper summits when compared with smooth

examples at the same nanometer scale and magnification. The parameter  $\Delta h_{\text{p-p}}$  is an effective measure of overall smoothness, especially since it can be correlated visually with an AFM scan represented as a 3-D plot. It is important to note that the comparison of different films on the basis of  $\Delta h_{\text{p-p}}$  is most meaningful when the film surfaces have similar textures.

Ilhan's (2003) wavelet summit texture analysis (WSTA) comparison is a rigorous and sensitive surface characterization technique that addresses the topography in the vicinity of the summits. The WSTA technique is not sufficiently sensitive to detect differences in  $\Delta h_{\text{p-p}}$ ; it is only sensitive to topography texture. Consequently, WSTA comparison is only suitable for texture comparison of films with similar  $\Delta h_{\text{p-p}}$ . For this reason we rely on  $\Delta h_{\text{p-p}}$  to compare results among various different experiments.

All scans reported here are 1000 nm on each lateral dimension, covering  $1 \mu\text{m}^2$ , comparable to the typical width of a transistor in modern CMOS technologies. The vertical axes represents the surface topography, and is greatly magnified as compared to the horizontal axes.

All film thicknesses reported here are measured with a Rudolph AutoEL laser spectroscopic ellipsometer, which contains calibrated routines for estimating the thickness of  $\text{SiO}_2$  films over a Si wafer, and of Si films in a Si– $\text{SiO}_2$  film stack over a Si wafer. Measurements using this ellipsometer are derived from changes in reflected polarization induced by thin films of known dielectric properties. Together with the type of the film stack, the angle of the incidence, and the thickness of the substrate film, ellipsometer observations are used to calculate the thickness of the film. The value used for the index of refraction is based on the available data for thicker films. The thickness of the underlying  $\text{SiO}_2$  substrate films is also measured with the same ellipsometer, using monitor wafers. Ellipsometric measurements for Si films with thicknesses  $>50$  nm have been verified against reflectance-based measurements using a NanoSpec optical spectrophotometer.

Note that the accuracy of the ellipsometric thickness measurements here is not known, primarily because of the unavailability of an effective value for the index of

refraction for very thin *a*-Si films with relative surface roughness. The ellipsometer manufacturer specifies a general figure of  $\pm 0.3$  nm for thickness accuracy of uniform films; no derating information for the film type or roughness is available, however. The effects of roughness, we note, are averaged over the area of the laser-beam spot. It is, therefore, important to note that all the film thicknesses reported here are effective ‘ellipsometric thicknesses.’

#### 4.2.2 Bare Si wafer surfaces

Si wafer vendors polish wafers to have atomically smooth surfaces, rendering these surfaces an ideal standard for this work. An oxide film a few angstroms thick, known as the ‘native oxide,’ forms rapidly on the bare Si surfaces when they are exposed to the atmosphere.

Figure 4.1(a) shows a surface scan of a bare Si wafer. At this magnification, the wafer surface is flat—even in the presence of native oxide. Note that  $\Delta h_{\text{p-p}}$  of this surface is 1.1 nm. Please note that Fig. 4.1(a) also demonstrates that our AFM system can resolve height variations at least as small as 1.1 nm,  $\Delta h_{\text{p-p}}$ .

#### 4.2.3 SiO<sub>2</sub> substrate surfaces

Figure 4.1(b) shows the surface of a Si wafer after it has been subjected to a 100-nm thick SiO<sub>2</sub> film growth at 850 °C in a wet ambient. The resulting wet-oxide surface has  $\Delta h_{\text{p-p}}=1.9$  nm, which is greater than that for a bare wafer surface but is still ‘smooth.’

In 3-D processing, the underlying SiO<sub>2</sub> films are LPCVD low-temperature oxide (LTO), as opposed to thermally grown wet-oxide or dry-oxide. For this reason we also examine LPCVD SiO<sub>2</sub> surfaces. Figure 4.1(c) shows the surface of a 25-nm thick LPCVD SiO<sub>2</sub> film deposited at 450 °C. Although only one-fourth the thickness of the films in Fig. 4.1(b), this LPCVD SiO<sub>2</sub> surface is significantly rougher, with  $\Delta h_{\text{p-p}}=4.3$  nm—a little more than twice that of thermally-grown SiO<sub>2</sub>. Similarly, Fig. 4.1(d) shows the surface of a 100-nm thick LPCVD SiO<sub>2</sub> film deposited at 400 °C

with  $\Delta h_{\text{p-p}}=4.7$  nm. There is no significant difference in  $\Delta h_{\text{p-p}}$  roughness between Figs. 4.1(c) and 4.1(d). This suggests that the  $\Delta h_{\text{p-p}}$  roughness of SiO<sub>2</sub> film reaches an equilibrium condition when the thickness is 25 nm or greater, although the summits have broadened somewhat in the 100-nm film. A comparison of the bare Si wafer surface of Fig. 4.1(a) and the 100-nm thick LPCVD SiO<sub>2</sub> surface of Fig. 4.1(d), however, reveals that the LPCVD SiO<sub>2</sub> surface is significantly rougher than the bare Si wafer surface. The roughness of the SiO<sub>2</sub> film surface impacts the surface roughness of films deposited over the SiO<sub>2</sub> substrate film.

#### 4.2.4 Thin Si film surfaces

##### Growth model

We use a simple model of CVD film growth, based on a random distribution of new nucleation growth sites along with growth of existing nuclei. In our model we assume that the substrate surface consists of a number of uniformly spaced potential nucleation sites, where each site is equally likely to initiate a nucleus. In this model, once a nucleate is formed it grows outwards hemispherically. Older nuclei eventually bury younger nuclei (Carter, 2000).

Figure 4.2(a) shows results from an example simulation of a 2-D deposition based on this model. This simulation indicates that the deposition process can be modeled as comprising three phases. In the first phase, A–B, the height variation  $\Delta h_{\text{p-p}}$  increases due to the formation and growth of nuclei that have not completely merged. In the second phase, B–C, the nuclei begin to merge, with the result that  $\Delta h_{\text{p-p}}$  decreases towards an apparent asymptotic value upon entering the third phase, C.

Our model is adequate for indicating qualitative trends. In general, we observe a similar trend in experimental studies of  $\Delta h_{\text{p-p}}$ , with the deposition progressing in a manner consistent with this model. The nucleation growth model does not take into account several important factors, however. In the model, the probability of starting a new nucleus is not affected by the presence of the existing neighboring nuclei. Claassen (1980) reports, however, that the nucleation density saturates with time.

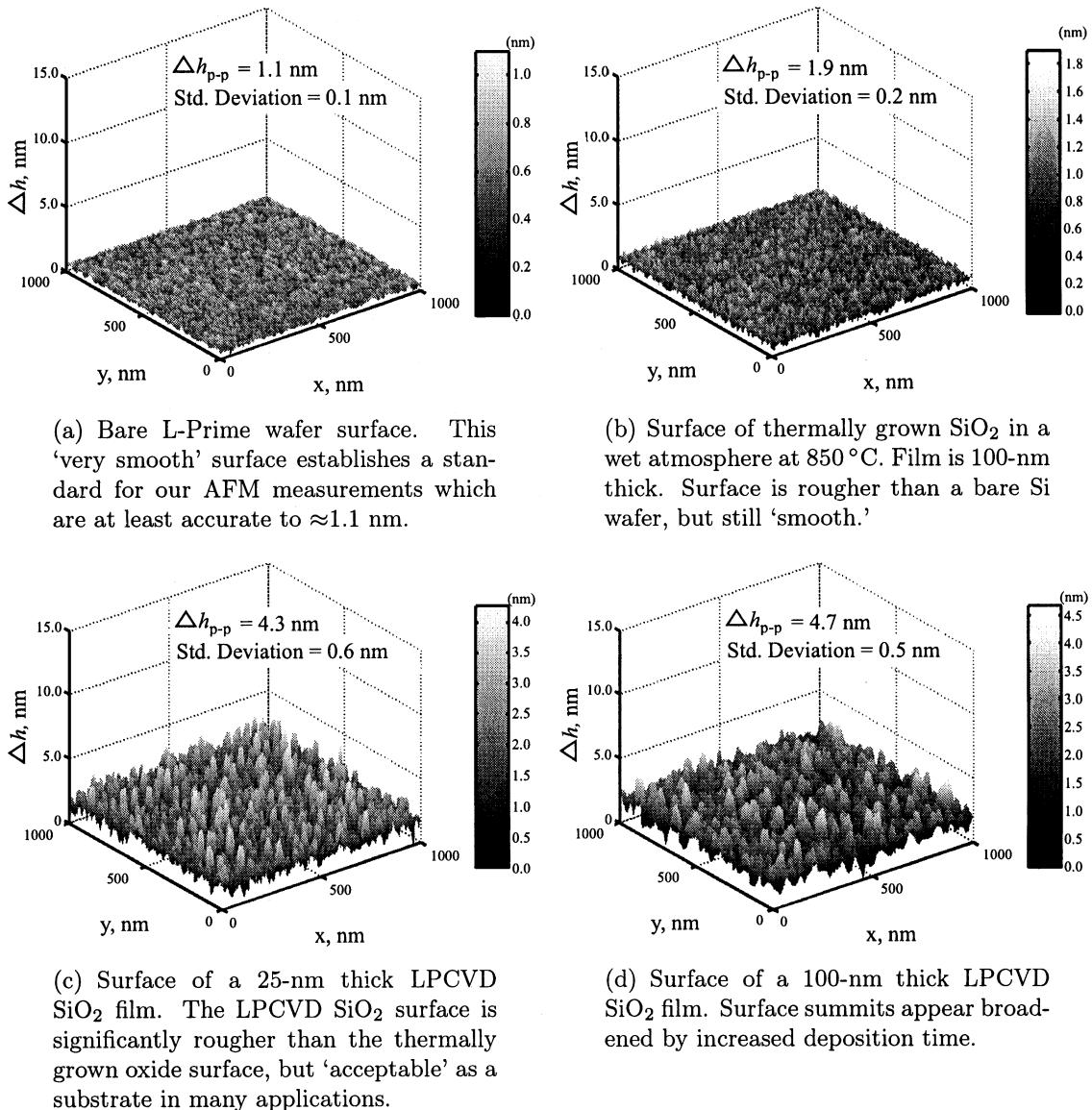


Figure 4.1: AFM scans of film surfaces.

As a result of this limitation, our model predicts a roughness peak earlier than when that actually occurs. In reality, the number of nuclei are fewer due to saturation, and the mean distance between nuclei increases. Longer deposition times are, therefore, required for nuclei to coalesce; hence, the roughness peaks later in time.

### **Effect of deposition process**

Figure 4.2(b) shows a surface scan of an *a*-Si film deposited for 3.75 minutes using a 560 °C LPCVD process at 400 mT<sup>1</sup>, with SiH<sub>4</sub> = 136 sccm<sup>2</sup> and H<sub>2</sub> = 110 sccm. A typical deposition rate for this process is 3 nm/min. Figure 4.2(b) shows the resulting tall summit texture on the surface, representing nuclei that have not yet merged—the film is not contiguous at this stage. Ellipsometer measurements indicate the thickness of this film to be around 9 nm. According to our deposition model of Fig. 4.2(a), this film is in the vicinity of point B.

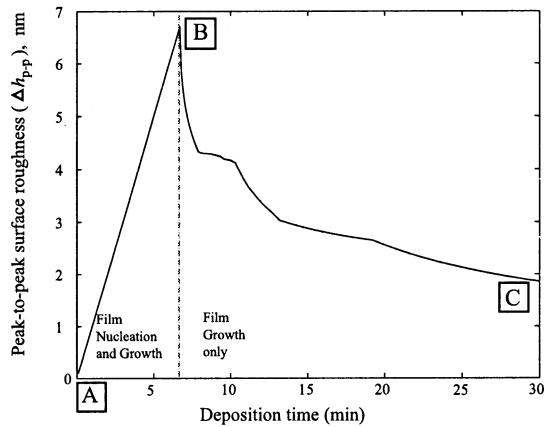
Figure 4.2(c) shows a similar film surface obtained using the same process, but after a longer deposition time of 24.0 min. At this time, the height variation  $\Delta h_{p-p}$  has decreased significantly after the longer deposition process, mainly because nuclei have merged together. Here, the deposited film thickness is about 70 nm.

A slower deposition process is more desirable for depositing films. An LPCVD *a*-Si process, with parameters 500 °C, 1000 mT, and SiH<sub>4</sub> = 200 sccm, yields a reduced deposition rate of 1 nm/min. Figure 4.3 shows surface scans of 9-, 18-, and 36-nm thick films deposited using this process. The film surface shown in Fig. 4.3(a) is in the early stage of nucleation and is rough. With increasing film thickness,  $\Delta h_{p-p}$  first increases for an 18-nm film surface as seen in Fig. 4.3(b), and then decreases for a 36-nm film surface as seen in Fig. 4.3(c). In Fig. 4.3(c), the nuclei have merged and so the deposited film is contiguous. This behavior is consistent with our model.

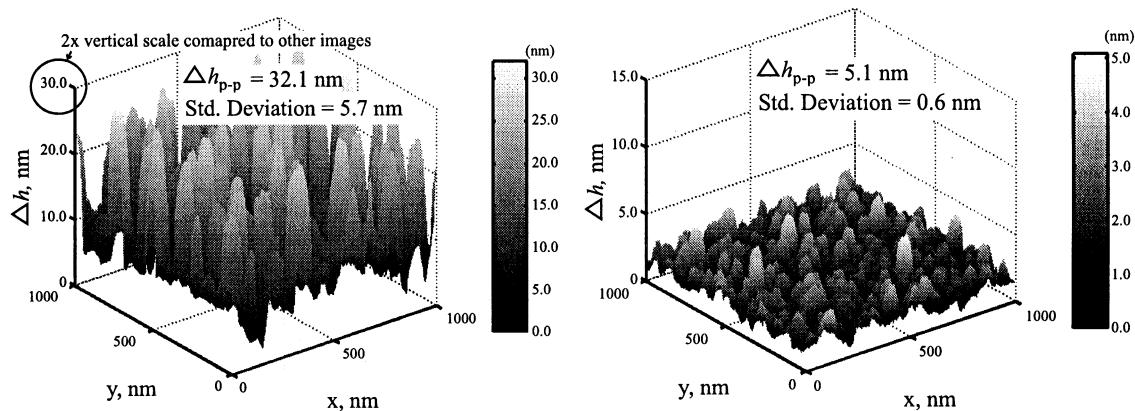
---

<sup>1</sup>In the Tylan LPCVD system used for this work, the pressure is controlled actively. A pressure controller, external to the LPCVD tube, injects nitrogen into a constant-speed vacuum pump, thereby stabilizing the system pressure.

<sup>2</sup>Standard cubic centimeters per minute (sccm).



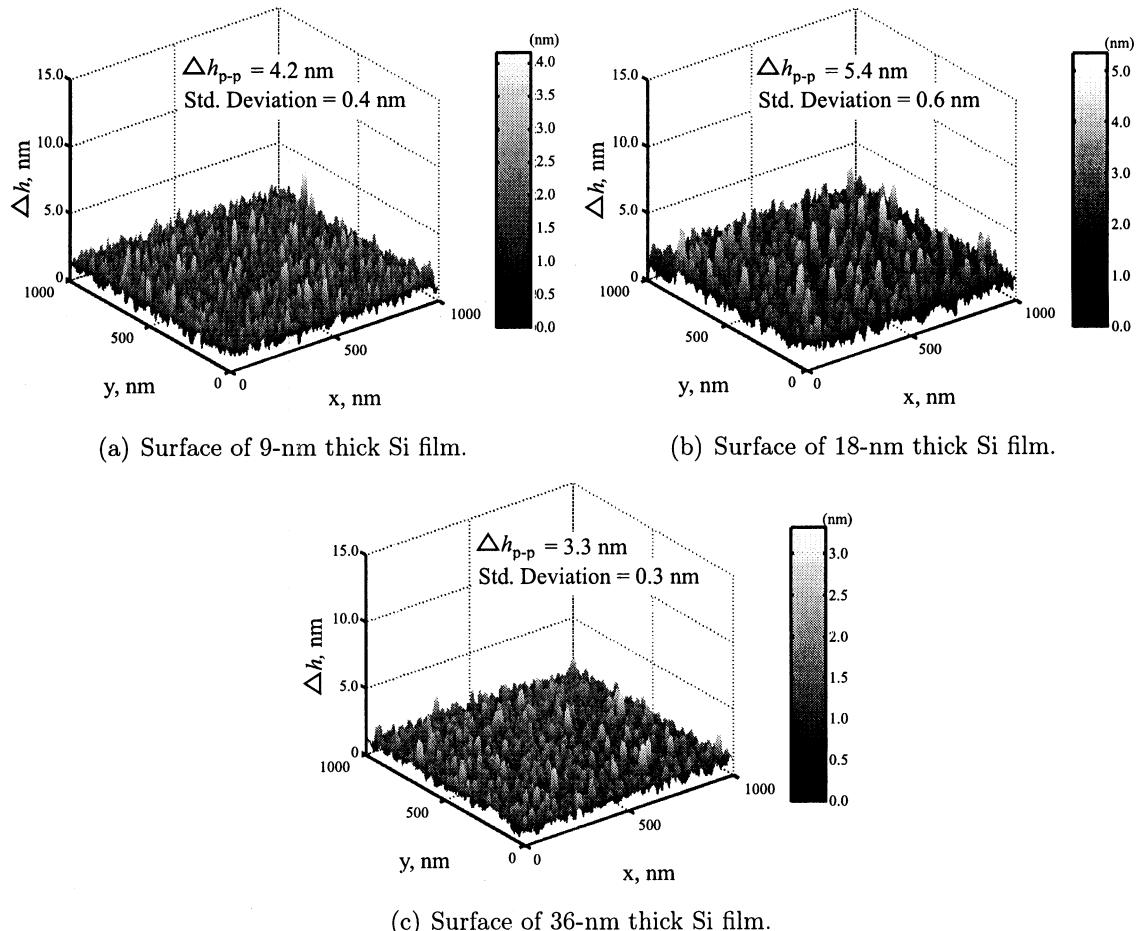
(a) Simulated 2-D surface roughness change with deposition progress. Nucleation starts at A, nuclei outgrow, and roughness increases until nuclei begin merging at B. Further film growth helps surface smoothness at C. Total deposited thickness = 30 nm; deposition rate = 1 nm/min; simulation step = 0.1 nm; nucleation probability/step = 0.1 %/site.



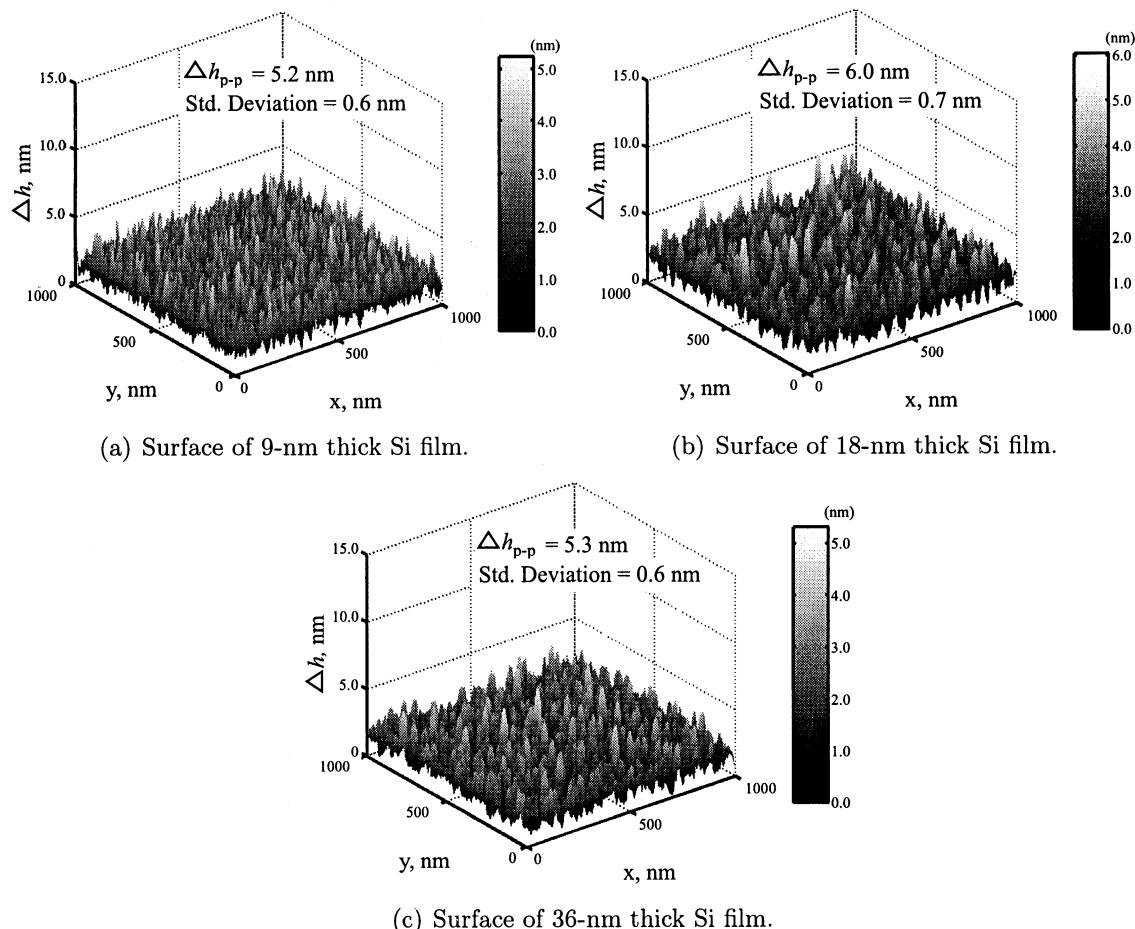
(b) Surface following 3.75 minutes of Si LPCVD deposition. Flow rates:  $\text{SiH}_4=136 \text{ sccm}$ ,  $\text{H}_2=110 \text{ sccm}$  at 400 mT, 560 °C. Note that the vertical axis is 2x compared with the other figures. This is a very fast deposition process; therefore, nucleation islands are large after only 3.75 minutes of deposition. Ellipsometric thickness is  $\approx 9 \text{ nm}$ .

(c) Surface following 24 minutes of Si LPCVD deposition. Flow rates:  $\text{SiH}_4=136 \text{ sccm}$ ,  $\text{H}_2=110 \text{ sccm}$  at 400 mT, 560 °C. Nuclei merge into a film after 24 minutes of deposition. Ellipsometric film thickness is  $\approx 70 \text{ nm}$ .

Figure 4.2: Progression of surface-film growth.



**Figure 4.3:** LPCVD Si deposition on 100-nm thermal oxide substrate. Deposition of Si is carried out at 500 °C, 1000 mT, and  $\text{SiH}_4 = 200 \text{ sccm}$ . Thickness values are from ellipsometry; see text.



**Figure 4.4:** LPCVD Si deposition progression on a 25-nm LPCVD SiO<sub>2</sub> substrate. Deposition of Si is carried out at 500 °C, 1000 mT, and SiH<sub>4</sub> = 200 sccm. Si film thicknesses are ellipsometric.

### **Effect of substrate roughness**

It may be seen in Fig. 4.1, by comparison of panel (b) with panel (c) and panel (d), that the LPCVD SiO<sub>2</sub> substrate surface is rougher than the surface of the thermal oxide. This roughness in the substrate may affect the roughness of the films deposited on top. Therefore it is important to examine closely the effects of substrate roughness on the surfaces of Si films deposited on oxide substrates. We carried out similar depositions of Si-film simultaneously on a 25-nm thick LPCVD SiO<sub>2</sub> substrate and on a 100-nm thick thermally grown oxide layer. The Si films shown in Fig. 4.4 have been deposited on a 25-nm LPCVD SiO<sub>2</sub> substrate. These films are similar to the previously described films of Fig. 4.3, except for the difference in the underlying oxide substrate.

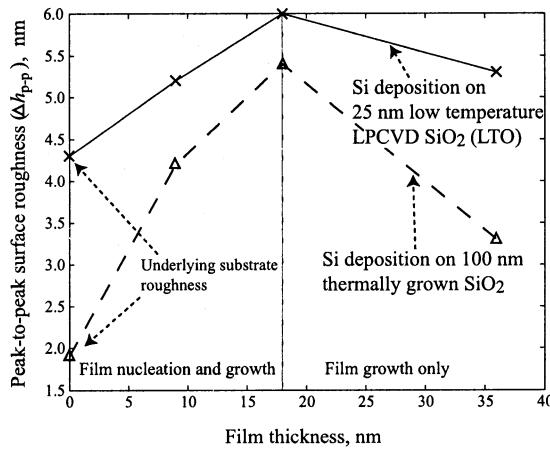
A comparison of films deposited on a smooth wet-oxide substrate vs. a rougher LPCVD SiO<sub>2</sub> substrate reveals noticeable difference in  $\Delta h_{p-p}$ , as shown in Fig. 4.5. It is clear that films deposited on rougher LPCVD SiO<sub>2</sub> substrates have comparatively higher peak-to-peak surface roughness. Especially for films thicker than about 36 nm, the difference in  $\Delta h_{p-p}$  is comparable, between the surfaces of the films deposited on the two kinds of underlying oxide substrates, to that between the surfaces of the underlying substrates themselves. We attribute this difference to the roughness of the underlying LPCVD SiO<sub>2</sub> substrates. Ilhan (2003) indicates that the surface summit texture is also different between the two surfaces.

### **Effect of precleaning**

Predeposition cleaning processes are a regular part of semiconductor fabrication. It has been shown that the grain size and surface roughness of LPCVD poly-Si films are affected by the surface-cleaning process (Voutsas & Hatalis, 1993). Our predeposition processing includes liquid-based cleaning with 4:1 H<sub>2</sub>SO<sub>4</sub>:H<sub>2</sub>O<sub>2</sub>, 5:1:1 H<sub>2</sub>O:H<sub>2</sub>O<sub>2</sub>:HCl, and 50:1 H<sub>2</sub>O:HF.<sup>3</sup> This cleaning process includes a 30-second immersion in the 50:1 H<sub>2</sub>O:HF solution that etches the SiO<sub>2</sub>. Moreover, X-ray fluorescence emission study of the cleaned wafer surfaces indicates the presence of S and F. Given this source of

---

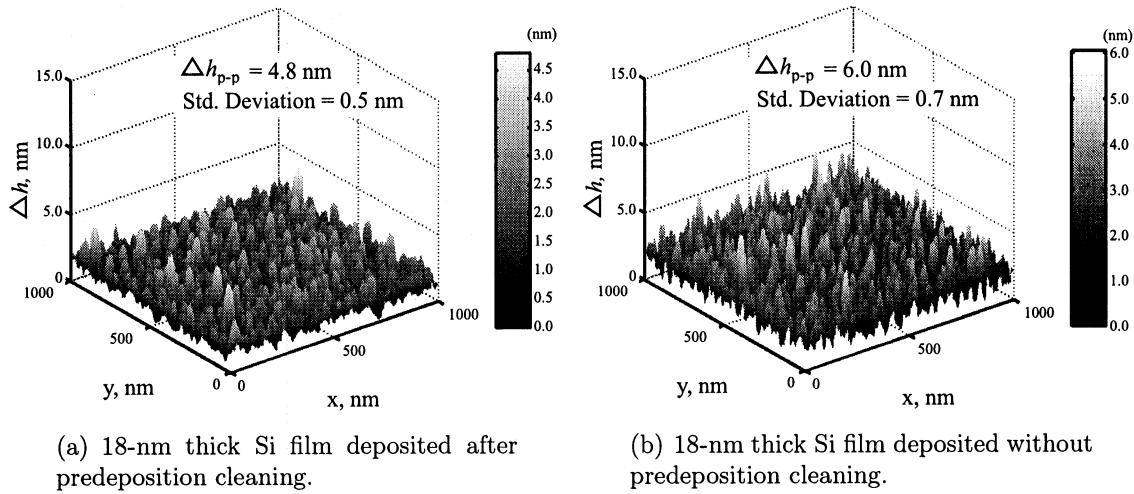
<sup>3</sup>All the cleaning mixtures are prepared according to volumetric ratios.



**Figure 4.5:** Peak-to-peak surface roughness ( $\Delta h_{p-p}$ ) of Si films on  $\text{SiO}_2$  substrate. Note that in both examples  $\Delta h_{p-p}$  increases until ellipsometric film thickness reaches  $\approx 18$  nm, and then decreases. This trend is consistent with our model, indicating that films thinner than  $\approx 18$  nm are in nucleation phase. Also,  $\Delta h_{p-p}$  is higher for films deposited on LPCVD  $\text{SiO}_2$  due to underlying LPCVD  $\text{SiO}_2$  surface roughness.

potential contamination, it is important to determine whether or not the presence of such impurities on the surface can change *a*-Si film surfaces as it does for poly-Si film surfaces.

In order to evaluate the impact of surface-cleaning treatments on surface roughness, we deposited similar films on similar substrates with and without a surface-cleaning treatment. A comparison of these film surfaces did not reveal significant roughness dependence on the predeposition cleaning. For example, Fig. 4.6 shows surfaces of an 18-nm Si film deposited on an LPCVD  $\text{SiO}_2$  substrate after a predeposition cleaning step, and without a cleaning step. The comparison shown represents the worst-case difference, measured after multiple scans. There is a difference of 1.2 nm in  $\Delta h_{p-p}$  between the surfaces in Figs. 4.6(b) and 4.6(a). As our AFM system can measure with confidence only down to  $\approx 1.1$  nm, as explained in Section 4.2, the difference of 1.2 nm is not sufficient to conclude that the cleaning process affects the *a*-Si surface roughness. The 1.2 nm, moreover, is a difference of two larger numbers, which could be the result of smoothing of the LPCVD  $\text{SiO}_2$  surface during the cleaning process. Further experiments are needed to resolve this question.

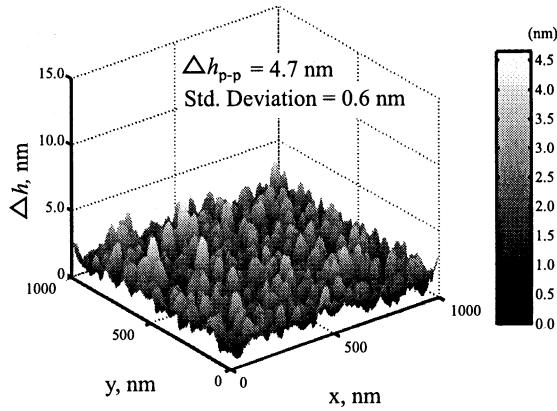


**Figure 4.6:** Surface of films on LPCVD SiO<sub>2</sub> substrate (a) with predeposition cleaning and (b) without any predeposition cleaning. Ellipsometric thickness of the underlying LPCVD SiO<sub>2</sub> substrate film is in the range of 25 to 28 nm.

Our conclusion regarding the effects of precleaning differs from that of Voutsas & Hatalis (1993). We attribute this difference to the use of different processes. Voutsas & Hatalis studied poly-Si on thermally grown SiO<sub>2</sub>, while here we address *a*-Si on LPCVD SiO<sub>2</sub>.

#### Etch-back smoothing

Figure 4.7 shows the surface of a 9-nm film deposited thickly and then etched back to an ellipsometric thickness of the layer in the range of 9 ~ 10 nm. This film was originally deposited with an estimated 25-nm thickness on an LTO substrate. A binary etch-back process using multiple cycles of surface oxidation and oxide etching was used to control etching to a specific thickness. The reduction of thickness by binary etching comprised alternate cycles of surface oxidation in HNO<sub>3</sub>, followed by removal of the oxidized surface film using a diluted HF solution at room temperature. Comparison of the reduced film with the surface of 9-nm thick film *a*-Si on LTO, shown in Fig. 4.4(a), indicates that the etched-back film surface has similar  $\Delta h_{p-p}$ , but a very different texture. The surface texture on the etched-back film has smoother summits, when compared with the film deposited with a thickness of 9 nm. This finding



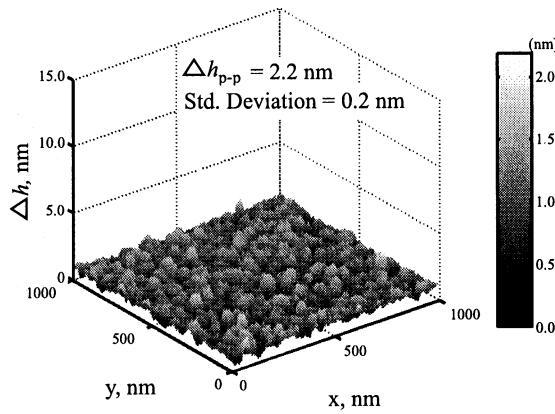
**Figure 4.7:** Effect of ‘etching back.’ This 9-nm thick film is etched back from an ellipsometric thickness of 25 nm.

indicates that ‘etching back’ of thickly deposited films is a reasonable approach to obtaining very thin films.

#### Effect of ion implantation and crystallization

Ion implantation and solid-phase crystallization also modify the film surface. To examine these effects, we use an LPCVD process at 560 °C, with 400 mT, SiH<sub>4</sub> = 136 sccm, and H<sub>2</sub> = 110 sccm, to deposit a 45-nm thick *a*-Si film on a base of 1000-nm thick thermally grown SiO<sub>2</sub>. Boron ions with 13-keV acceleration energy and an ion dose of  $1 \times 10^{15} \text{ cm}^{-2}$  are then directly implanted into the *a*-Si film. Subsequently, the *a*-Si film is crystallized using a two-step annealing process, in which the film is first exposed to 700 °C for 5 minutes in a rapid thermal annealing machine, after which it is placed in a furnace at 600 °C for 4 hours.

Figure 4.8 shows the surface of the *a*-Si film after implantation and annealing. This surface is significantly smoother than the film surface shown in Fig. 4.2(c), which is 70-nm thick (*a*-Si) and is not implanted. While implantation and crystallization processing smooth the surface, such processing requires exposure to temperatures higher than 600°C, which unfortunately renders this approach unsuitable for UTB applications. It would be interesting to investigate the impact of ion implanting Si into Si, however.



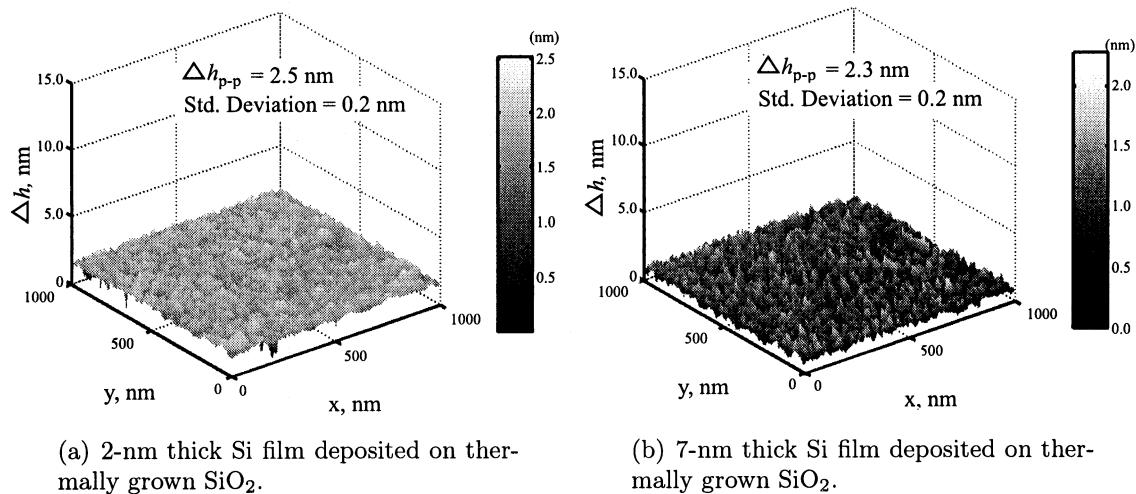
**Figure 4.8:** Surface of 45-nm thick Si film after ion implantation and crystallization.

Ion implantation with an ion dose on the order of  $10^{15} \text{ cm}^{-2}$  damages single-crystal Si, converting it into an amorphous state. Solid-phase crystallization, as is typically used in CMOS processing, reestablishes the crystallinity. A number of similar underlying phenomena may be responsible for surface smoothing. These include surface flattening by ion bombardment, surface restructuring as a result of heavy ion dosage, and solid-phase crystallization later in the process flow.

#### Surfaces of Si films deposited by source evaporation

Si films deposited by evaporation from a solid source are of particular interest because they can be deposited on a substrate at room temperature. Such physically vapor-deposited (PVD) films have been reported to be porous (Fritzsche & Tsai, 1979). Fulks *et al.* (1999) report that laser-crystallized TFT channels, using *a*-Si deposited at room temperature by evaporation, show only about one-half the mobility is observed with TFTs using LPCVD *a*-Si. In certain applications where near-room-temperature processing is the only option, for instance for silicon on flexible substrates, such mobility values achieved by evaporated films are still very attractive.

Our study of such *a*-Si films shows increased surface smoothness relative to CVD films. Figure 4.9(a) shows a 2-nm thick *a*-Si film deposited by means of evaporation. The film displays  $\Delta h_{\text{p-p}}=2.5 \text{ nm}$ , which is comparable to the thermal oxide surface presented in Fig. 4.1(b). Figure 4.9(b) shows 7-nm thick *a*-Si films deposited by a similar process. This film shows  $\Delta h_{\text{p-p}}=2.3 \text{ nm}$ , which is again comparable to that of



**Figure 4.9:** Surfaces of Si films deposited by evaporation of an Si source.

a thermal oxide surface. Very smooth surfaces of Si films deposited by evaporation are as expected. Intuitively, the use of lower temperatures decreases the probability that the surface adsorbing species will diffuse on the surface, thereby leading to smoother films.

*Now that we have seen some ways to keep surface roughness at bay, we next turn our attention to controlling lithography at very small feature sizes, in order to create tiny, power-thrifty circuits. All the better, we now show that we can accomplish this objective with existing equipment and well-understood present-day standard operating procedures.*

## Chapter 5

### Edge-Defined Lithography

To reach the next level of reduced feature sizes, we need to “think outside the box.” Difficulties are increasing with traditional approaches of projecting mask images onto silicon surfaces, as feature sizes become comparable to the wavelength of the source of illumination. But again, we can “push the envelope” dramatically, using the technique of edge-defined lithography.

Lithography, the process of producing patterns on a surface, is a prime semiconductor processing step, that must be optimized in order to shrink the dimensions of transistor channels. Minimizing device gate lengths is important for ultra-low-power (ULP) technology because transistors with such shorter channel lengths can deliver higher or comparable driver currents for lower supply voltages than transistors with longer channel lengths. Therefore, ULP circuits using transistors having shorter channel lengths operate at higher frequencies, for comparable power dissipation, than circuits using transistors having longer channel lengths.

Successful implementation of ULP technology requires the use of lithographic techniques that produce nanometer-scale channel lengths in large volume and at practical cost. Of particular interest is the development of 3-D ULP technology, which requires a lithographic technique that can produce nanometer-scale channel lengths in 3-D stacked circuits while making use of a low-temperature processes only, because the temperature cannot be allowed to exceed  $\approx 600^{\circ}\text{C}$  for long durations after forming the first level of devices. This limitation on the processing temperature

results from the need to avoid undesirable redistribution of dopant profiles in the lower-layer devices that have been already created.

Lithographic techniques based on extreme ultraviolet (EUV) light, optics-immersed-in-liquid, x-rays, electron beams, and scanning probes all show promise as methods that can produce patterns with resolutions in the nanometer-scale range. EUV, and projection lithography combined with immersion optics, are the most promising techniques at the 65-nm CMOS technology ‘node’ and beyond. Both systems are, however, expected to remain expensive even after their technical problems are solved. Mask-set prices for 130-nm technology and beyond are in the \$1,000,000 range today. This price level essentially prohibits small-volume, commercial production.

Edge-defined patterning, sometimes referred to as ‘spacer-defined’ or ‘deposition-defined’ patterning, has the potential to provide a low-cost, high-throughput, and well-controlled lithographic approach, while extending the life of existing photoresist and projection lithography systems. The basics of edge-defined techniques are well understood, having been studied for over two decades in various contexts (Okazaki *et al.*, 1981; Hunter *et al.*, 1981; Hunter *et al.*, 1981; Flanders & Efremow, 1983; Malhi *et al.*, 1984; Takasu & Todokoro, 1984; Strifler & Cantos, 1990; Horstmann *et al.*, 1998; To & Woo, 1999; Zhang *et al.*, 2001; Choi *et al.*, 2002).

In this chapter we focus on the development of an edge-defined, nanometer-scale patterning technique suitable for low-thermal-budget 3-D IC fabrication. First we describe different types of lithographic techniques and summarize their features. Next, we describe edge-defined lithography, processing flow, and design rules. We then present and analyze the experimental details and results of edge-defined lithography experiments carried out as part of this research. Finally, we show transistor layout techniques for edge-defined lithography.

## 5.1 Lithography

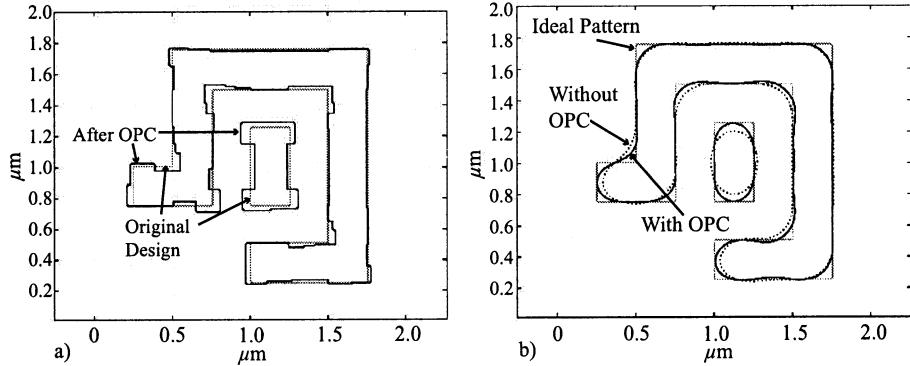
Semiconductor lithography is used to place patterns on a crystalline wafer—a process often referred to as patterning. In application, a photosensitive material, commonly referred to as photoresist, is coated over the wafer. This coating is usually applied to a spinning wafer so that the rotational speed controls the photoresist thickness. The desired patterns are applied to the photoresist by exposure to light or to an electron beam, changing the chemical solubility of the photoresist in the exposed areas. Exposed or unexposed photoresist is washed off the wafer after reaction with a chemical developer. For a positive photoresist system, it is the exposed photoresist that is dissolved by the developer. For a negative photoresist system, it is the unexposed photoresist that dissolves.

### 5.1.1 Lithographic techniques

Lithographic processes are divided into two main categories, mask-dependent and maskless. The mask-based approach is more prevalent in mass production, but mask prices are high because of the need for labor-intensive processes used to achieve resolution enhancement. On the other hand, the maskless approach is ideal for prototyping, although throughput is small—a few wafers per hour. Indirect edge-defined lithography makes use of a conventional mask, and achieves finer patterning resolution than optical resolution through the use of precisely controlled deposition thicknesses. We now describe these approaches in more detail.

#### Mask-based lithography

Mask-based lithography requires masks containing all final feature details. The main advantage of a mask-based lithographic system is its high production rate, about one wafer every minute. A typical mask is created on a scale with a pattern around five times larger than the final pattern. Optical reduction of the mask image then shrinks the features to their intended size. The limit to final feature resolution depends mainly on the wavelength of the light used to transfer the mask image onto the wafer.



**Figure 5.1:** Example improvements from OPC. a) Mask design, before and after OPC. Note that the OPC serifs are used to emphasize the corners. b) Simulated intensity contours resulting from the original mask, the OPC mask, and the desired feature. The intensity contour represents the shape of the pattern achieved with that particular mask. Clearly the mask with OPC optimization achieves a pattern closer to that desired as compared with the pattern achieved with the original mask.

Fabrication of a “perfect” mask is key to a successful manufacturing process. All production masks undergo inspection, and every defect has to be discovered and rectified before the mask is used for production. With millions of transistors on a single chip, production of a defect-free mask set is a labor-intensive and costly process. The considerable cost of the masks makes any revision or design change expensive once a mask set has been prepared.

In order to extend projection lithography to nanometer-scale feature sizes, optical optimizations are used. Formation of sharp corners, bends, and critical dimensions are the most difficult aspects of projection photolithography scaling. A technique called optical proximity correction (OPC) is usually used to improve bends and corners. Figure 5.1 shows an example of OPC for a mask feature (Cobb, 1998). The OPC technique divides each polygon into small segments, and adds resolution assist features to each segment. Applying OPC to the mask layout improves resolution, process reliability, and yield.

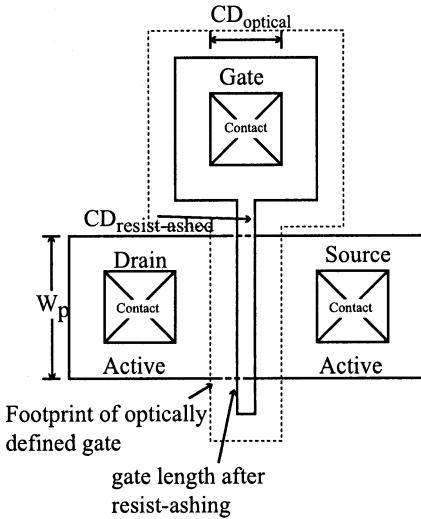
While OPC has helped to extend the life of projection lithography, there are undesirable effects associated with this process. When applied to complex mask

designs, OPC processing introduces directional distortions resulting in severe local variations.

Mask-based lithographic resolution depends on the projection optics as well as on the photoresist system resolution. With conventional mask and optical systems, the finest resolution achievable is about one half to one third of the wavelength of the illumination. With advanced optical processing of the mask, features as small as one third of the wavelength are routinely achieved. For example for 130-, 90-, and 65-nm technology feature sizes, lithographic systems are based on 193-nm wavelength laser-light sources. It is likely that 193-nm projection lithography combined with immersion optics will allow projection lithography to achieve feature sizes smaller than  $\approx 45$  nm.

Electron projection lithography (EPL) is yet another possibility, as is 13-nm-light-source-based extreme ultraviolet (EUV) lithographic technology. The commercial viability of EUV technology has not been established to date. EUV masks consist of complex mask blanks comprising multilayer thin-film stacks. These films are typically only a few nanometers thick. Production of defect-free mask blanks, repair of EUV masks, and the availability of continuous high-power light sources with a cost-effective lifespan are major concerns for widespread adoption of EUV systems.

In addition to optical optimization, photoresist etching techniques are used to further reduce the minimum critical dimension achievable with optics. Figure 5.2 shows the footprint of a transistor formed with these methods. In this approach, first a longer gate is defined, and then it is uniformly reduced in all dimensions. This technique, commonly known as “photoresist-ashing” or simply “resist-ashing,” is applied only to gate-patterning lithography. It consists of patterning the photoresist using projection lithography, followed by an oxygen-plasma-based oxidizing process for controlled etching of the photoresist to reduce feature sizes. The term ‘ashing’ is, therefore, used as opposed to ‘etching,’ because the controlled oxidation of the photoresist is chemically the same as burning it. The resulting uniformity of the final features is, however, rather poor. Another limitation of the resist-ashing process is



**Figure 5.2:** Photoresist ashing. In submicron CMOS technologies, optical lithography assisted by resist-ashing is used to reduce the minimum dimension.

that photoresist etching can only shrink block-type features; it cannot shrink trench-type or contact-hole-type features.

### Maskless lithography

Maskless processes effectively store the mask in the software. The motivation for the maskless approach is clear: the cost and complexity of mask production and handling are greatly reduced, simplifying the lithographic processing. These techniques suffer from a very low production throughput, however, and are applied mainly to the production of masks for prototypes. Maskless systems based on electron-beam and laser sources are now used for preparing prototypes. Scanning probes and EUV-based maskless systems have been proposed and are being researched.

#### 5.1.2 Lithography selection criterion

In order to understand the value of edge-defined lithography, we need to first understand the important aspects of a lithographic system. A brief overview of the important lithographic parameters follows.

### Throughput

Commercial viability of lithographic systems usually is coupled to their throughput. The throughput of a lithographic system depends upon the underlying lithography technique employed, and is affected by downtime and by servicing needs. Some lithographic systems are inherently slow. For example, e-beam lithography exposure of a single wafer can take hours. Thus, such systems are suitable only for research purposes.

Conventional projection lithography usually is combined with an automated loading and exposure platform, commonly known as a stepper. The stepper greatly enhances throughput by automatically aligning the reticle to the existing patterns on every die, focusing and exposing the photoresist through a mask, and then stepping to the next die. Steppers can process about one wafer every minute.

### Reliability and yield

As in other industrial semiconductor production processes, lithography must not only be reliable, but also must achieve high yields. Controlling of process parameters and minimizing defects are usually the main concerns. As an example, in the early stages of 193-nm immersion lithography development, the defect density introduced by the flowing liquid was high. Control of such defects will ultimately determine the manufacturing suitability of any lithography system.

### Cost of mask fabrication

Over the past few years there has been a remarkable increase in the cost of masks, as the need for precision has increased. The cost of a 1- $\mu\text{m}$  mask set prepared during the year 2000 for research purposes at the Stanford nanofabrication lab was only a few thousand dollars. Today, for state-of-the-art 90-nm technology, the cost of a mask set is over a million dollars. Economic pressures due to mask price leaves little room for experimentation with circuit designs, especially for small businesses. For high-volume production, the one-time high mask cost can be easily absorbed in overhead. Low-volume production, however, is severely limited by high mask costs.

### Ease of mask inspection, repair, and handling

It is important to discover and repair every mask defect prior to its use in device fabrication. A single mask defect, if not corrected, can result in complete production failure. Handling and storage of masks is also an important consideration. For example, EUV masks require storage and handling in a particle-free atmosphere with controlled temperature and humidity. This requirement complicates the storage and handling of EUV masks, adding to the complexity of using this system.

## 5.2 Edge-Defined Lithography

Our research requires a low-cost, repeatable, high-throughput lithographic system for nanometer-scale devices. Edge-defined lithography offers a combination of the necessary attributes. Potentially, it could extend the resolution of any lithographic tool to the next generation. The complexity of the mask is not increased, although the required processing time does increase, however, due to required additional deposition and etching steps. The throughput of a commercial edge-defined processes might be comparable to that of conventional photolithography.

In this process, we first define a spacer around a sacrificial registration block. The sacrificial block is then stripped away, leaving the outline of the registration in the form of the spacer. The spacer width is defined by the spacer-film thickness, and controls the final feature width by acting as a mask. A second pattern can then be superimposed at this stage with a patterned photoresist, enabling fabrication of edge-defined and photo-defined features at the same level. The photoresist and spacers are then used to pattern the final features.

### 5.2.1 Processing details

The process described here relies on a dual-material system—chemically vapor-deposited silicon, and silicon dioxide. This system is particularly attractive because HBr-plasma-based-etching processes etch silicon and have good selectivity between

silicon films and silicon dioxide films. Similarly, HF-based-wet-etching processes etch silicon dioxide and not silicon. Availability of these etching processes makes it possible to selectively etch one material and stop the etching process when the other material is exposed, and enables the use of dual material for this process.

### Spacer formation and patterning

Figure 5.3 illustrates the main steps of our low-temperature edge-defined lithography—fabrication of an edge-defined feature, and its integration with a photo-defined feature.

In Fig. 5.3(a), the Si film to be patterned is coated with an ‘etch-stop’ layer, comprising a deposited layer of low-temperature LPCVD  $\text{SiO}_2$ . This  $\text{SiO}_2$  deposition is carried out at  $450^\circ\text{C}$ , 250 mT, 5 sccm<sup>1</sup>  $\text{SiH}_4$ , and 70 sccm  $\text{O}_2$ , which gives a very slow deposition rate of roughly 3 nm/min. The etch-stop layer is designed to absorb etching and deposition variations in the spacer and the registration films.

Figure 5.3(a) also shows the registration layer, which is a layer of silicon deposited with LPCVD at  $620^\circ\text{C}$ , 400 mT, 136 sccm  $\text{SiH}_4$ , and 110 sccm  $\text{H}_2$ . In another set of experiments, this layer was deposited as LPCVD amorphous silicon at  $550^\circ\text{C}$ , 400 mT, 136 sccm  $\text{SiH}_4$ , 110 sccm  $\text{H}_2$ . This layer is patterned to provide the sidewall edge that is used for the final feature definition.

Figures 5.3(b) and 5.3(c) illustrate patterning of the registration layer. This step uses photolithography to define registration polygons, illustrated here by a simple rectangular box over the registration layer, Fig. 5.3(b). The dimensions and spacing of the polygons depend upon the limits of the photolithographic technology employed. In addition, the line-edge roughness (LER) of the registration process limits the smoothness of the patterned feature. Registration polygons are etched in silicon, Fig. 5.3(c). The etching process uses 100 sccm  $\text{HBr}$ , 5 sccm  $\text{O}_2$ , 40 sccm  $\text{Cl}_2$ , 250 W top RF power, and 60 W bottom RF power at 10 mT, after which the photoresist is removed.

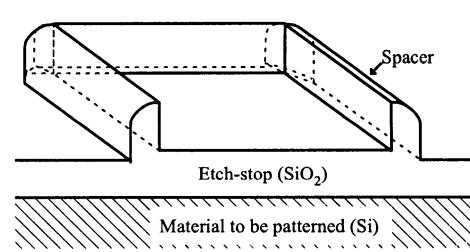
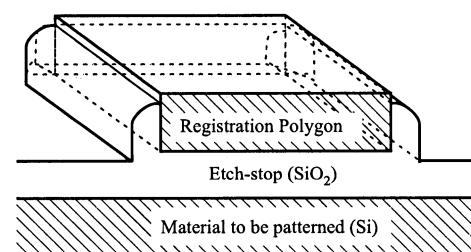
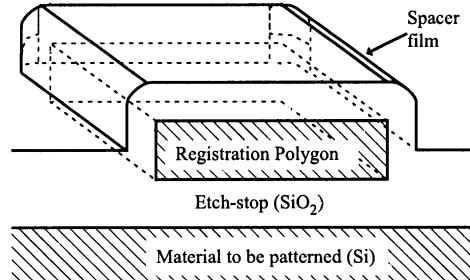
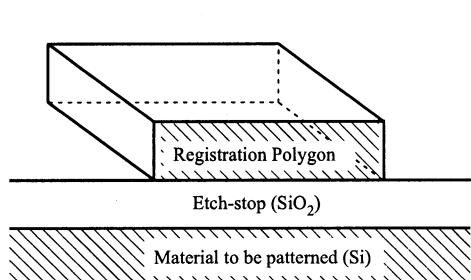
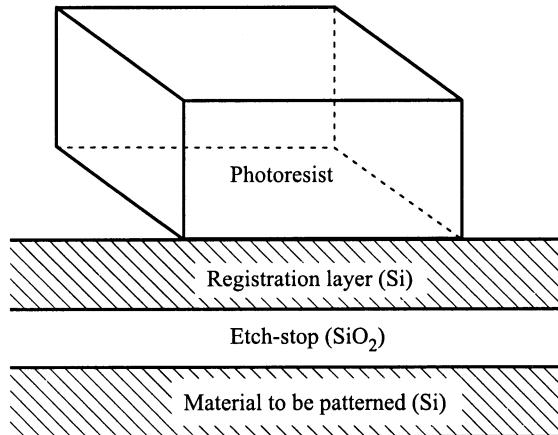
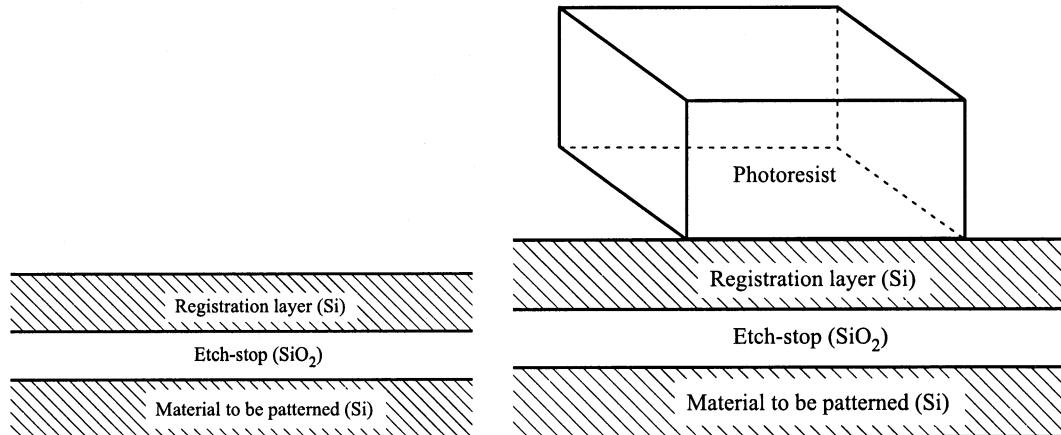
Next, a second LPVCD  $\text{SiO}_2$  film, covering the registration polygons, is deposited in preparation for forming a ‘spacer,’ as shown in Fig. 5.3(d). Here, the thickness of

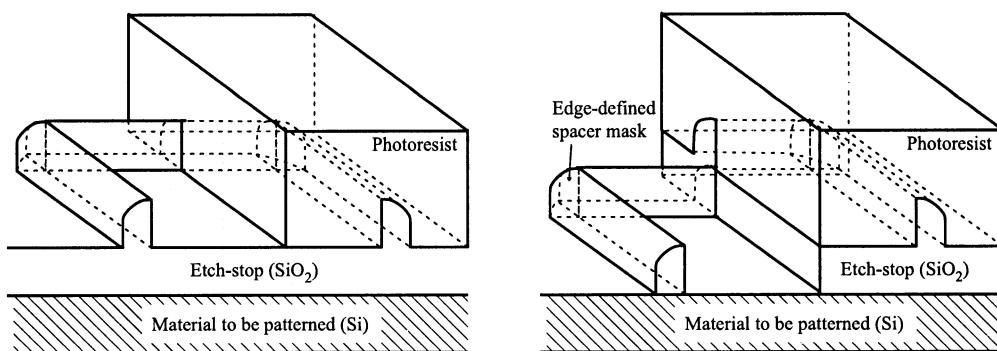
---

<sup>1</sup>Standard cubic centimeters per minute.

*Next two pages.*

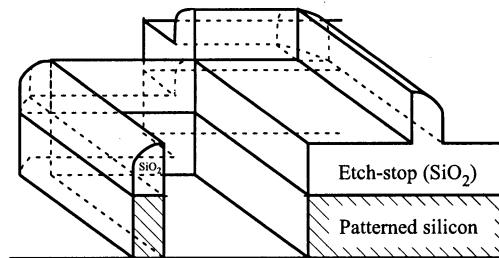
**Figure 5.3:** Edge-defined patterning process flow.





(g) Integration with second lithography step.

(h) Edge-defined spacer oxide anisotropic etch-back



(i) After final patterning etch, photoresist removed.

this film controls the final structure width. The spacer is formed by etching silicon dioxide in an Applied Materials AME8100 etcher using 85 sccm CHF<sub>3</sub>, 6 sccm O<sub>2</sub>, 400 mT, 1400 W RF power, and electrode bias of -530 V. This process results in anisotropic etching, and in polymer buildup that protects the sidewall from lateral etching and undercutting. The spacer appears as the sidewall surrounding the registration polygon, as seen in Fig. 5.3(e).

After spacer formation, the remaining silicon in registration polygons is removed using 50 sccm HBr, 5 sccm O<sub>2</sub>, 250 W top power, and 45 W bottom power in the poly-etcher; see Fig. 5.3(f). This process gives about 10:1 etch selectivity between silicon and silicon dioxide. Monitoring the silicon emission spectrum of the etching byproducts generates a clear endpoint signal, which is used to stop the etching process.

Photoresist is spun on and optically patterned to define contact pads and other structures, as suggested by Fig. 5.3(g). This step integrates optically- and edge-defined patterns at the same lithographic level. Silicon dioxide is etched away to remove the underlying silicon using the previously mentioned anisotropic dry etch, leaving a hard mask of silicon dioxide for the edge-defined portion and a photoresist mask for the photo-defined portion, Fig. 5.3(h). A Lam Research silicon etcher is then used to create the final pattern in silicon; see Fig. 5.3(i).

It is possible to use other types of etch processing for patterning different materials. The most critical part of the gate etch is controlling undercut while obtaining high selectivity to oxide, in order to stop the etch on thin oxide films in gate patterning. Endpoint detection on the Lam TCP 9400 enables the gate etch to be stopped on thin gate oxide.

### Spacer removal

In certain applications, it may be necessary to remove the SiO<sub>2</sub> spacer mask after patterning with edge-defined lithography, for instance, to expose gate Si for silicidation during transistor fabrication. We now describe a second processing flow that can be used to remove the spacer mask.

Figure 5.4 shows the process for removing the oxide spacer mask while patterning the final features. After the formation of the oxide spacer hard mask, the final feature is not completely etched into the silicon, Fig. 5.4(a). A thin blanket silicon film remains on the wafers. The wafers are then subjected to a diluted HF solution, and the oxide hard mask is stripped away, leaving the structure shown in Fig. 5.4(b). The blanket silicon film protects the underlying oxide films from dissolving during this HF solution. Wafer are then subjected to silicon etching process in order to remove the thin blanket silicon film.

Figure 5.4(c) shows the final features obtained after spacer removal.

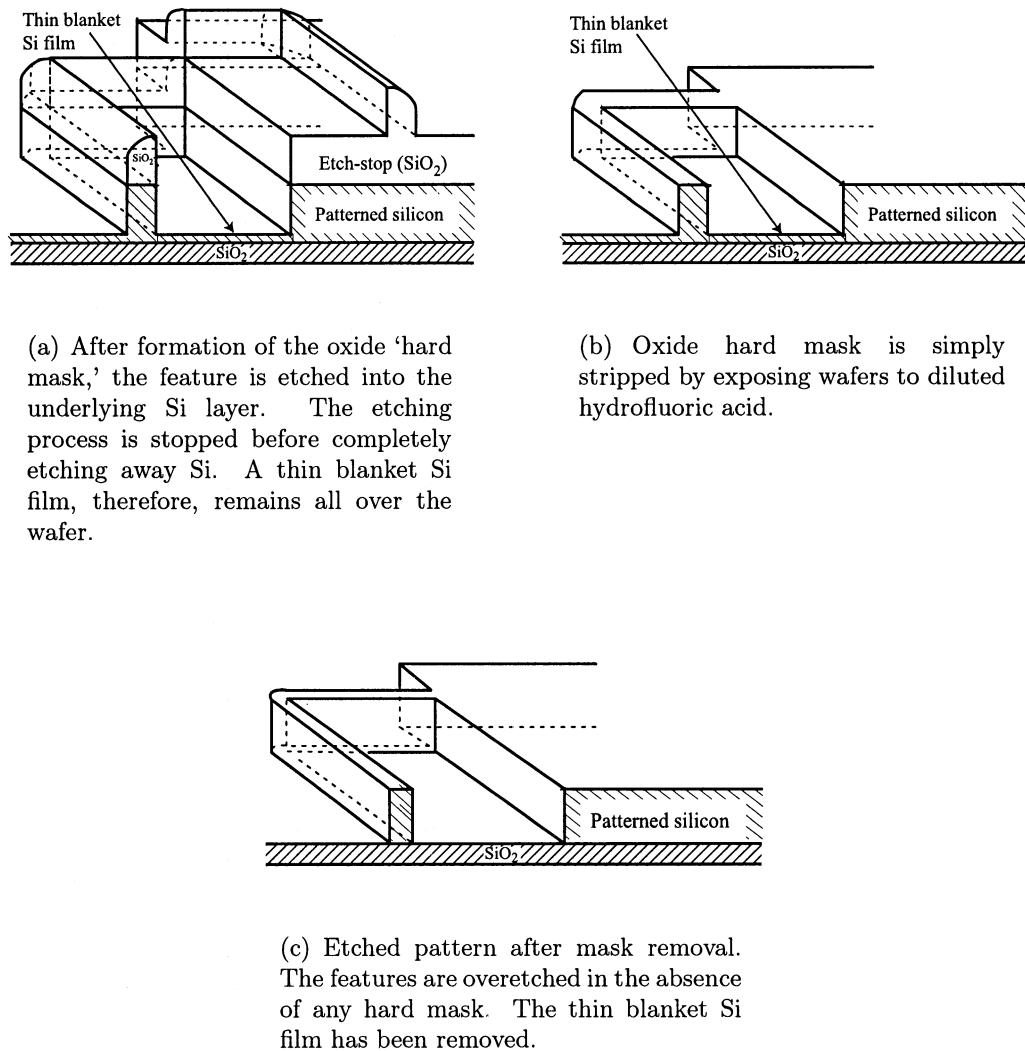
## 5.3 Experimental Results

One objective of our edge-defined patterning experiments is to develop a technology for MOS transistor gate structure formation. Our following experiments, therefore, are used to pattern structures similar to transistor gates. Silicon films, patterned in these experiments using edge-defined lithography, have been deposited over a thin  $\text{SiO}_2$  film that we refer to as gate-oxide film. This gate-oxide film is 2–4 nm thick LPCVD  $\text{SiO}_2$ , which represents a realistic thickness for an MOS-transistor gate dielectric.

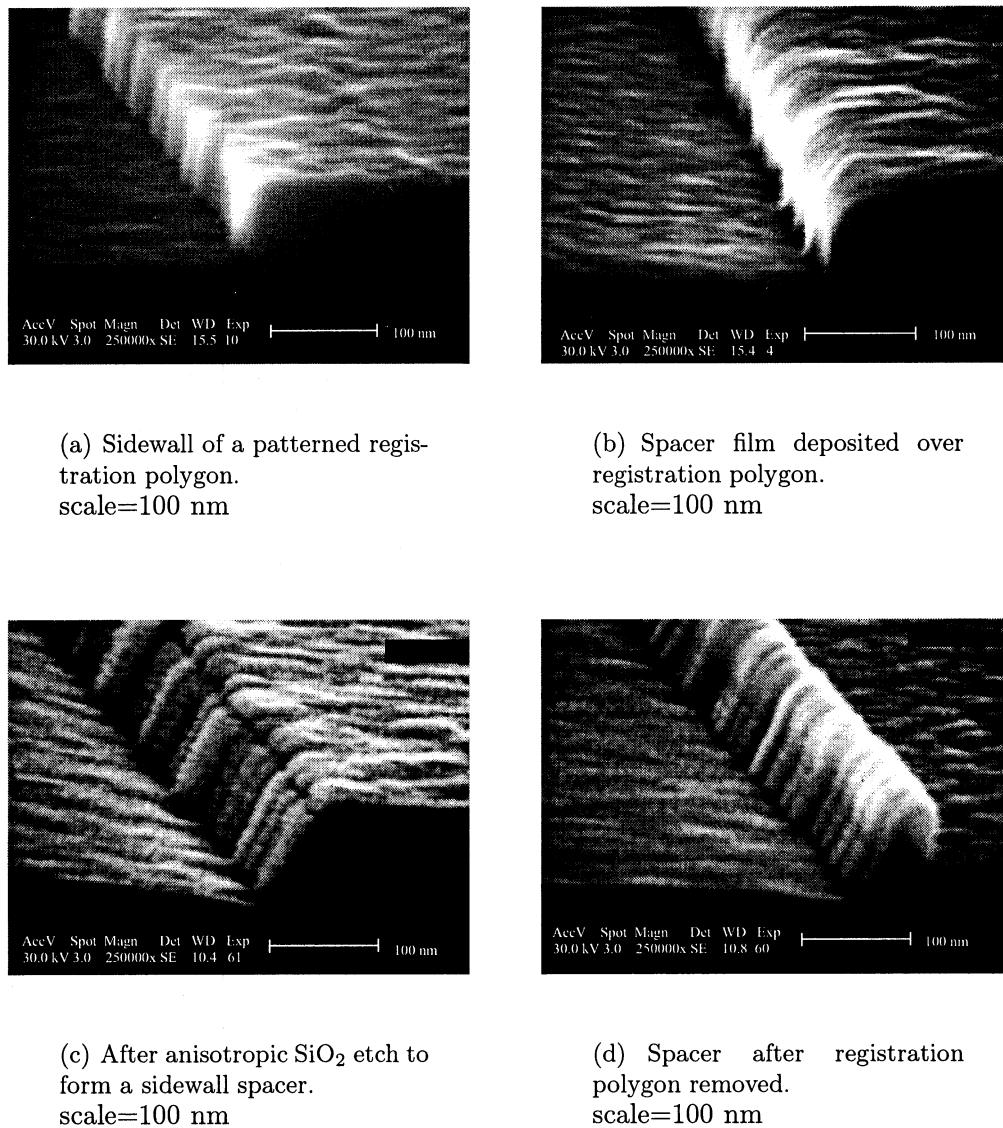
### 5.3.1 SEM images during edge-defined patterning

Figure 5.5 shows a series of SEM images taken after different processing steps in a 50-nm width edge-defined patterning experiment. It is clear from Fig. 5.5(a) that the registration polygon sidewall is not straight. The undulation in the sidewall of the registration polygon is about 10 nm. This variation results, we believe, from our lithographic process which is optimized for a 1- $\mu\text{m}$  feature size, and silicon etching for edge registration.

These undulations do not affect the smallest features, because the final pattern conforms to the edge. Figure 5.5(b) shows that the LPCVD  $\text{SiO}_2$  spacer film has a



**Figure 5.4:** Edge-defined patterning-spacer mask removal process. This processing flow is an alternative to the processing shown in Fig. 5.3(i). Note that these figures also show the underlying  $\text{SiO}_2$  film.



**Figure 5.5:** SEM images for an edge-defined patterning experiment.

similar rough nodular structure. Undensified LPCVD SiO<sub>2</sub> and phospho-silicate glass (PSG) films have a porous structure. Figure 5.5(c) shows the result of anisotropic etching of the LPCVD SiO<sub>2</sub> film to make a spacer. Figure 5.5(d) shows the spacer after the registration silicon has been removed. It is clear that the spacer is rough, and this roughness transfers to the next layers when the spacer is used as etching mask.

### 5.3.2 Critical dimension and line-edge roughness

Critical dimension (CD) and line-edge roughness (LER) are general terms used to describe the quality of a patterned structure, as illustrated in Fig. 5.6. The CD is the smallest manufacturable feature size that a lithography system can achieve. Figure 5.6(a) defines the critical dimension. In deep sub-micron technologies, there are significant variations in CD across each wafer and even within each die. CD variations are defined in Fig. 5.6(b) and (c). In Fig. 5.6(b) two opposite sidewalls of a pattern do not follow a straight line due to processing variations; an interesting case of sidewall variation is shown in Fig. 5.6(c), where the two sidewalls track each other so that the width of the feature remains constant. In the latter case, it is important to note that there are variations along each single sidewall. The variations along the sidewall are referred to as line-edge roughness (LER), as shown in Fig. 5.6(d).

### 5.3.3 Sample preparation

Samples for SEM imaging were prepared by cleaving the Si dice along a dense array of edge-defined structures. The cleaving process results in the dice breaking along the straight crystallographic orientation. This procedure produces silicon dust that can be blown away with nitrogen gas; some residue can still be seen in the images, though. High-resolution imaging was, for the most part, obtained with a 2 kV electron-beam acceleration. Some of the photomicrographs were obtained using 25 kV electron-beam acceleration, however, which is relatively higher than the optimal energy for

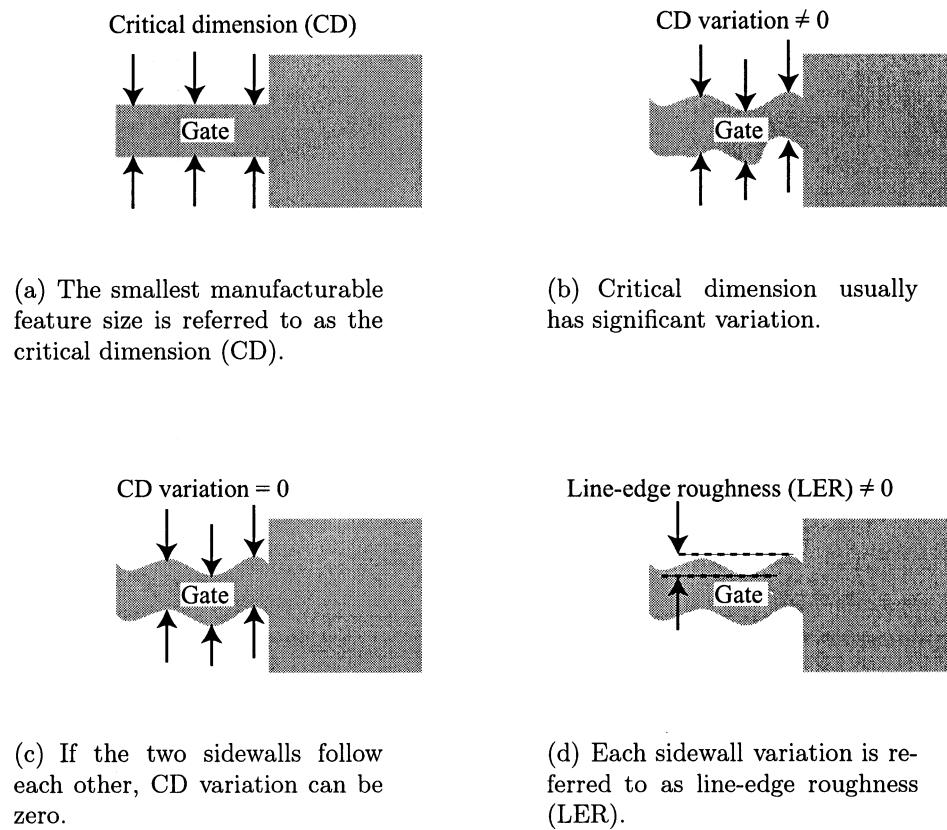


Figure 5.6: Definitions of CD and LER.

fine resolution. This was necessary for proper machine operation at the time of imaging.

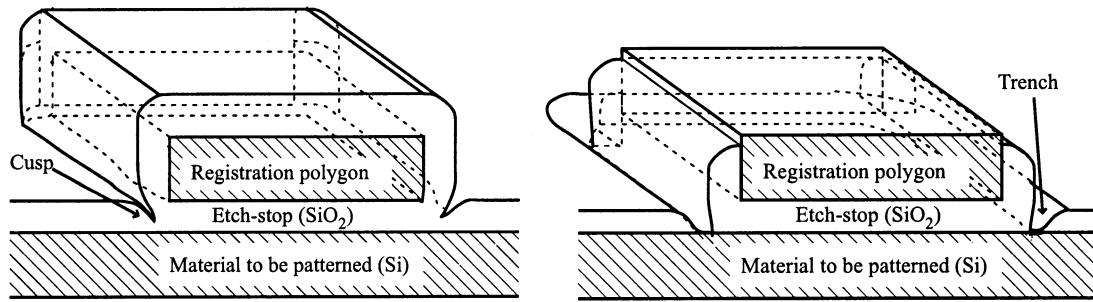
### 5.3.4 Undesirable trench formation

Most of the edge-defined lithographic techniques presented in the literature use at least three distinct films in the process: silicon, its oxides, and its nitrides. Typically the third material, such as nitride, is very dense, and is deposited at high temperatures that are not tolerable in a 3-D process. The process here is simpler in that it uses only silicon and its oxide; both can be deposited at low temperatures, and are well-characterized semiconductor materials.

This simplification of the process depends on the  $\text{SiO}_2$  etch-stop film that is deposited, cf. Fig. 5.3(a)–(d). Design of this film thickness is critical for the process. If this film is too thin, it can result in trenching of the underlying Si layer; if it is too thick, it can make it difficult to control the etching processes due to large relative variations in the etching of thicker films.

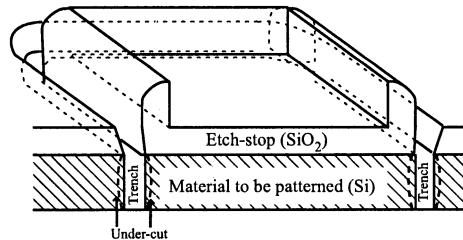
The trenching problem is explained in Fig. 5.7. Figure 5.7(a) shows a registration block with an  $\text{SiO}_2$  spacer film deposited over it.  $\text{SiO}_2$  deposition processes do not necessarily yield a conformal spacer film. A cusp can appear at the concave corners, in which case the deposited thickness is usually less near the corners (Kao *et al.*, 1985). Figure 5.7(b) suggests how a trench can form around the spacer during the spacer-formation etching process. As a result, due to the thinly deposited oxide at the concave corners, the etch-stop oxide may be consumed completely during the etch processing. When this occurs, the underlying film is exposed.

Figure 5.7(c) shows an undesired deep trench formed around the spacer. Here, the etching process used to strip away the registration silicon also removed silicon from the underlying film, resulting in the failure of the process. Figure 5.8 is a photomicrograph of an 180-nm edge-defined spacer with a trench on the right side. The stop-etch oxide film thickness was only 20 nm in this experiment. Proper film thicknesses are,



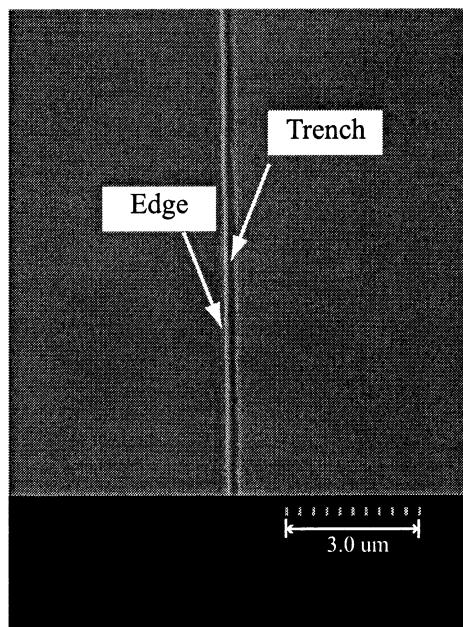
(a) LPCVD  $\text{SiO}_2$  deposition is not very conformal as a spacer film. A cusp shows up around the corners, and the deposited thickness is usually less at the corners.

(b) During the spacer formation etching process, thinly deposited  $\text{SiO}_2$  can result in a trench around the spacer. If the etch-stop  $\text{SiO}_2$  film is not sufficiently thick, the underlying film will be exposed.



(c) If a deep trench exists around the spacer, the etching process used to strip the registration silicon can also etch away silicon from the underlying film. This results in the failure of the process.

**Figure 5.7:** Trenching problem in the edge-defined patterning flow.



**Figure 5.8:** 180-nm edge-defined pattern with an undesired trench on the right side.  
scale=3  $\mu\text{m}$

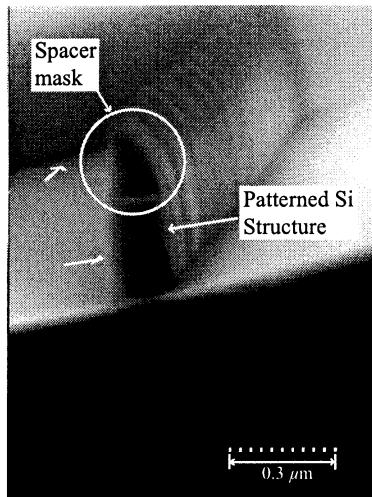
therefore, critical to avoid the trench problem and to maintain control of the process. Design rules to achieve process optimization are presented in a subsequent section.

### 5.3.5 180-nm edge-defined patterning

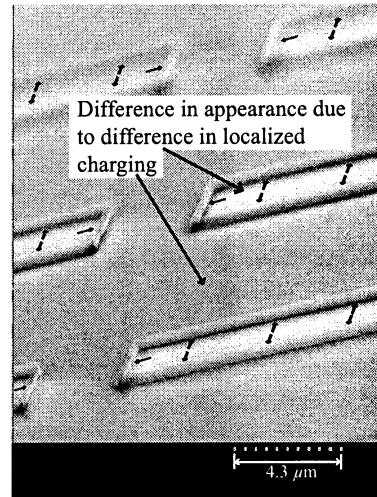
Figure 5.9 shows the results of a 180-nm width edge-defined patterning experiment.

A closeup cross-section of a patterned structure is shown in Fig. 5.9(a). The oxide spacer cap is clearly visible on each structure. A careful look at the sidewall of the spacer reveals vertical streaks in the sidewall. These streaks comprise sidewall roughness resulting from processing after spacer-film deposition, and contribute to CD variations. Figure 5.9(b) shows the edge-defined structures formed along the perimeters of registration polygons; arrows point to location of spacers.

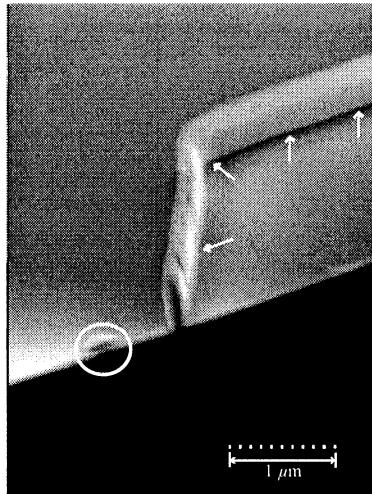
Figure 5.9(c) shows a closeup of a 180-nm wide edge-defined structure. The encircled object is a particle resulting from the sample preparation procedure. The white arrow points to the location of the perimeter of the registration polygon before it was removed.



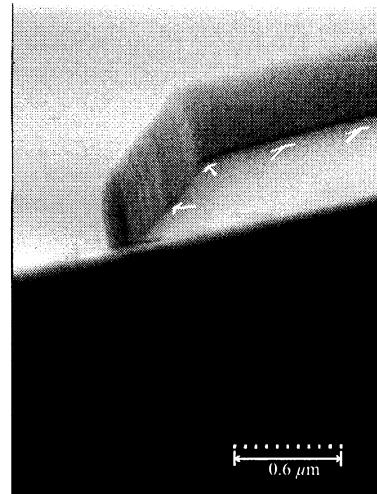
(a) Edge-defined 180-nm structure. The  $\text{SiO}_2$  spacer mask is visible on top of the pattern. scale=0.3  $\mu\text{m}$



(b) Edge-defined patterns along the perimeters of the registration polygons that have been removed. scale=4.3  $\mu\text{m}$



(c) Edge-defined 180-nm structure. The arrows show the previous location of the boundary of the registration polygon. The encircled object is a silicon dust particle. scale=1.0  $\mu\text{m}$



(d) edge-defined 180-nm structure. Note the roughness on the spacer sidewall formed next to the previous location of the registration polygon. scale=0.6  $\mu\text{m}$

**Figure 5.9:** 180-nm edge-defined lithography results. The arrows show the previous location of the boundary of the registration polygon.

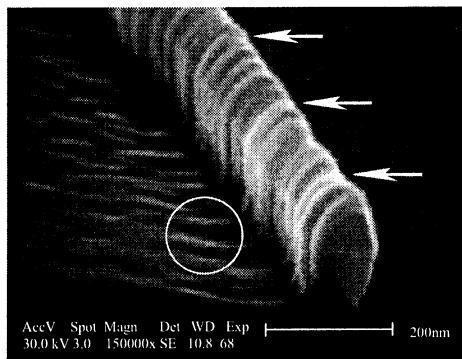
Figure 5.9(d) shows a closer image of the same structure, but from a different angle. Note that streaking is also visible on the sidewall facing the registration polygon. These streaks result from the non-uniform sidewall of the registration edge. The sources of this roughness are the processes used to pattern the registration film.

### 5.3.6 90-nm edge-defined patterning

Figure 5.10 shows patterned edge-defined structure with widths of 90 nm. The spacer cap is clearly visible on top of the structure. The white arrows indicate the previous locations of the sidewalls of the registration polygons. The circle indicates surface damage resulting from the final etching process that created the gate. This can lead to process difficulties when this edge-defined process is not properly used.

As mentioned earlier, the patterned Si film is deposited over a LPCVD SiO<sub>2</sub> film in order to mimic the MOS gate structure. This underlying LPCVD SiO<sub>2</sub> film is only 2-nm thick in this experiment. The final etching processing, therefore, has to be stopped before it breaks through this thin film of LPCVD SiO<sub>2</sub>.

Si etching processes used here have a nominal 10:1 etching rate selectivity between Si and SiO<sub>2</sub>. Note that this Si etching process does etch SiO<sub>2</sub>, albeit relatively slowly. In the areas where registration polygons were located previously, the thickness of the Si film to be patterned is greater than that of the Si film outside the registration polygons. Therefore, during the final Si patterning etching process, the underlying thin LPCVD SiO<sub>2</sub> film gets exposed earlier in the area outside the registration polygons, and has to withstand longer exposure to the Si etching process. The additional exposure of this thin SiO<sub>2</sub> film to a slow Si etching process can result in excessive removal of SiO<sub>2</sub> in areas where the film is thin or defective, eventually resulting in pit formation as seen in the figure. Such etching damage has to be carefully taken into account when using edge-defined techniques for transistor fabrication.



**Figure 5.10:** 90-nm edge-defined patterned structure. The white arrow points to the registration polygon side. The white circle points out surface damage, resulting from the Si etching process that broke through the thin gate dielectric  $\text{SiO}_2$  film. Note: horizontal scale is 200 nm.

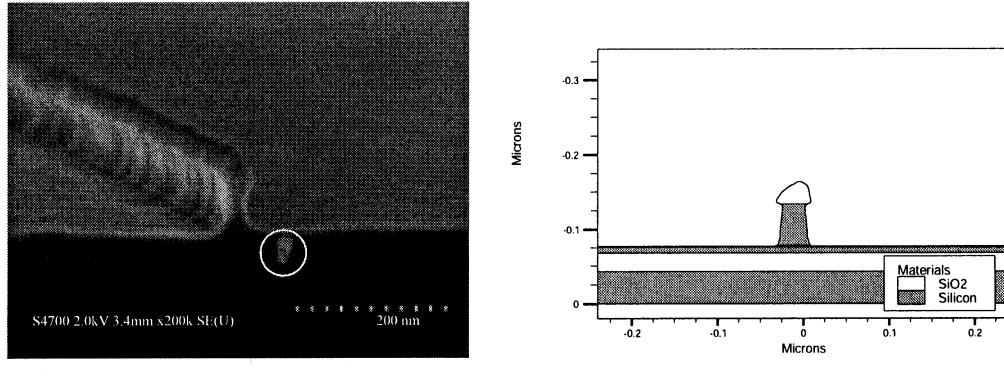
We also note that the edge-defined structure shows streaks or ‘caterpillar-like segments.’ This CD variation is a consequence of the uneven edge-registration sidewall, the uneven spacer film, and etching micro-variations.

### 5.3.7 $\leq$ 50-nm edge-defined patterning

Figure 5.11(a) shows the result for a 50-nm wide structure, patterned with edge-defined lithography. This figure shows a perspective view of a patterned structure with an oxide spacer on top. There is significant undercutting; note the concave sidewall resulting from the etching process. Figure 5.11(b) shows a simulated cross-section profile of a similar structure, which predicts the actual structure very well.

Figure 5.12 is a SEM micrograph of a 50-nm wide structure. The ‘caterpillar-like’ sections of the structure are clearly visible.

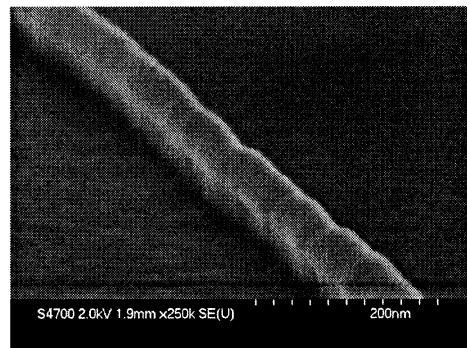
Figure 5.13 shows a 25-nm edge-defined structure without a spacer mask cap. The spacer mask is removed using the process described previously. Note that the white arrows show the previous location of the registration polygon. Surface damage due to etching is also visible in the areas outside of the previous location of the registration polygon, for the same reason described earlier.



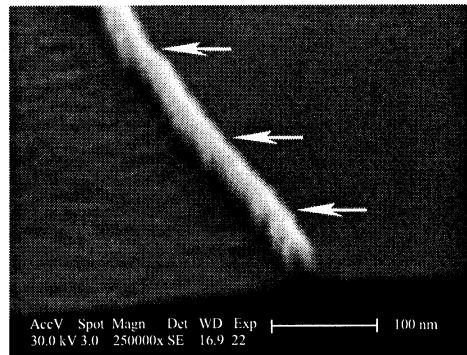
(a) An SEM micrograph showing an aerial view of a 50-nm edge-defined patterned structure. The oxide spacer mask is at the top of the structure. There is significant roughness along the patterned structure. The encircled object is a dust particle.

(b) The cross-section profile resulting from process simulation. The undercut in the final feature is due to poor etch selectivity in the final patterning process.

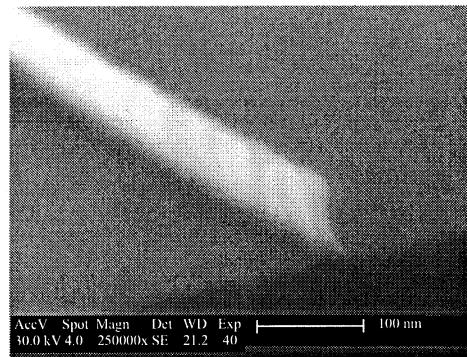
**Figure 5.11:** 50-nm edge-defined patterned structure.



**Figure 5.12:** Side view of a 50-nm structure with an oxide spacer mask. The CD variation resulting from the caterpillar-like structures is obvious.



**Figure 5.13:** 25-nm structure without oxide mask. scale=100 nm.



**Figure 5.14:** 18-nm edge-defined structure. scale=100 nm.

Figure 5.14 shows an 18-nm edge-defined structure. In this particular experiment, the oxide spacer has not been removed; therefore the structure appears taller than it actually is, as compared to the one seen in Figure 5.13.

### 5.3.8 CD-SEM analysis

Choi *et al.* (2002) report long-range average CD control better than 7 nm using spacer definition. Their process employed a high-temperature oxide for the etch-stop layer, which is incompatible with 3-D IC processing. Our process results in short-range CD variation that is roughly the same as that reported by Choi *et al.* (2002) for e-beam lithography. As for long-range CD variations, LPCVD processes can be controlled to within a few percent. Across-the-wafer, wafer-to-wafer, and lot-to-lot CD variations

are a function of LPCVD process control. It seems reasonable then that such CD variation can, therefore, be controlled to a few percent as well.

Figure 5.15(a) shows CD-SEM analysis of our 50-nm structure. It is clear that the short-range CD variation along a 200-nm section is in the range of 4 to 8 nm. Similarly, Figs. 5.15(b), 5.15(c), and 5.15(d) show additional CD-SEM measurements. CD-SEM analysis indicates about the same standard deviation of critical dimension ( $CD-\sigma$ ) for 50-, 25-, and 18-nm wide structures.

Patterned structures show a caterpillar-like, segmented morphology (cf. Fig. 5.11(a)). This morphology is attributed to i) registration-edge roughness, ii) polymer formation during the oxide dry-etch process, and iii) LPCVD SiO<sub>2</sub> spacer-film roughness. The polymer results from C–F based etch chemistry, producing rough sidewalls by acting as a micro-etch mask. The spacer LPCVD SiO<sub>2</sub> film surface itself shows nodular structure under AFM. An LPCVD SiO<sub>2</sub> film with a thickness of 28 nm shows surface roughness of  $\sigma = 1$  nm, and a range  $\approx 2\text{--}4$  nm p-p. Understanding of silicon dioxide and silicon film surface roughness is important to control CD variations for nanoscale features.

The CD-SEM analysis was carried out on a Hitachi S-4700 SEM that can resolve 2.1 nm at 1 kV. We were careful to minimize any distortion in the SEM images. It is still possible, however, that some of the jagged edges resulted from charge-up or e-beam damage during SEM data acquisition.

### 5.3.9 Design rules

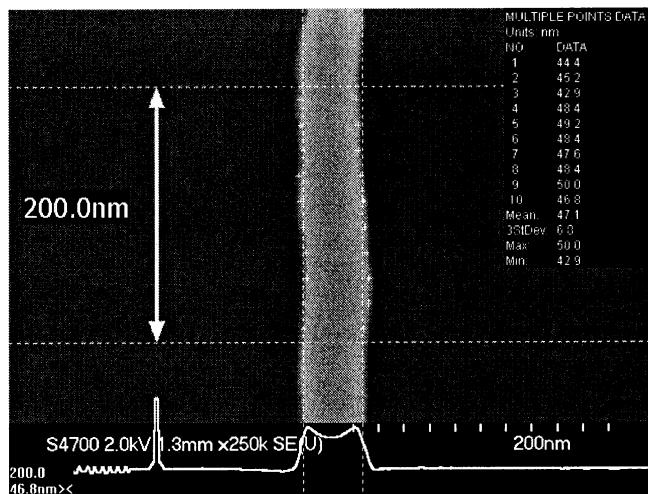
Based on our experiments, we formulated the following design rules, which we used several times to repeat the process. Figure 5.16 defines the process-critical film thicknesses.

For  $L_g > 90$  nm, the process depends mainly upon the proper choice of film-stack thickness. For a target  $L_g$ , the film stack thicknesses are as follows:

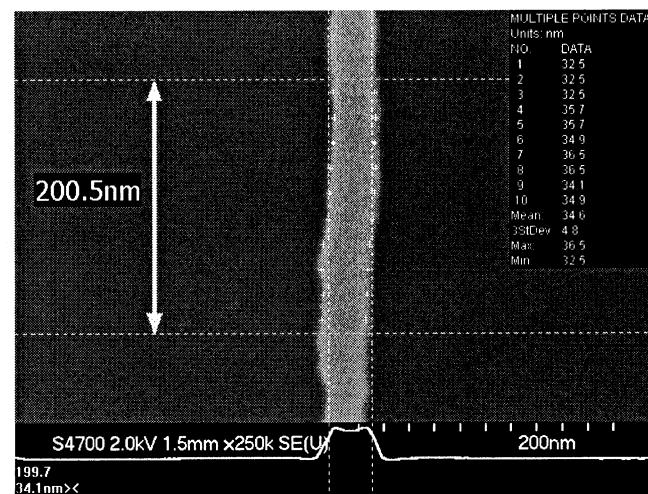
$$Z_{sp} \approx L_g$$

*Next two pages.*

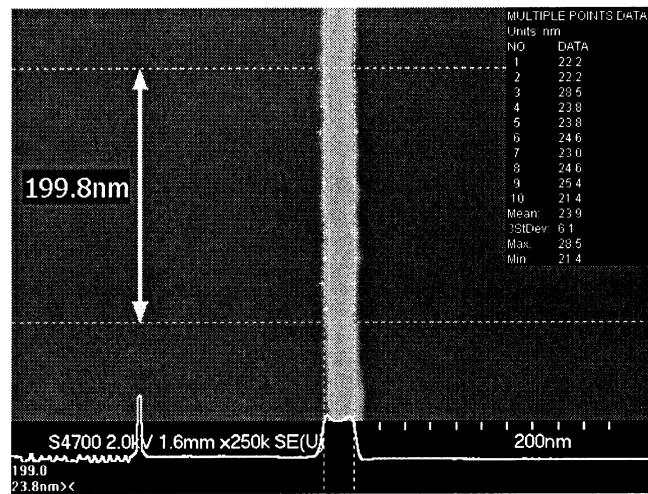
**Figure 5.15:** CD-SEM analysis of edge defined structures. This analysis is carried out on sample structures from different parts of a wafer for each experiment. The analysis software measures the distance between the boundaries of the patterned structure as viewed from the top of the die. Lower trace represents intensity of the image along the lower dotted line.



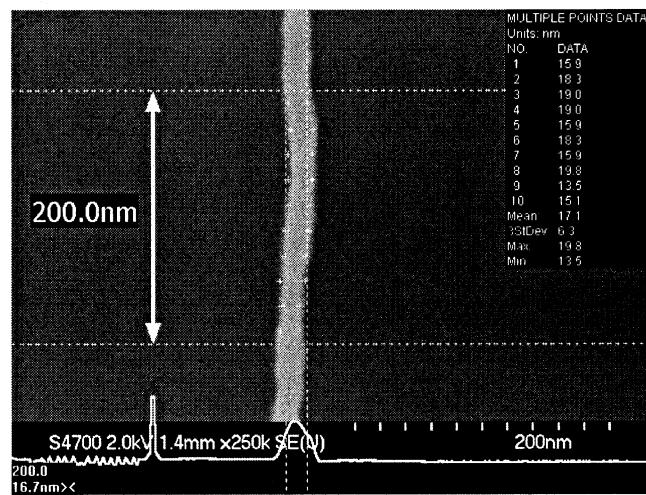
(a) 50-nm edge-defined structure.  
scale=200 nm mean=47nm  $\sigma$ =6.3 nm  
Range=50.8–42.8nm



(b) 35-nm edge-defined structure.  
scale=200 nm mean=34.6 nm  $\sigma$ =4.8 nm  
Range=36.5–32.5 nm



(c) 25-nm edge-defined structure.  
scale=200 nm mean=23.9 nm  $\sigma$ =6.1 nm  
Range=28.5–21.4 nm



(d) 18-nm edge-defined structure.  
scale=200 nm mean=17.1 nm  $\sigma$ =6.3 nm  
Range=19.8–13.5 nm

$$Z_r \approx 1.8 \times L_g$$

$$Z_{st} \approx 1.4 \times L_g$$

However, for edge-defined structures below 90 nm and above 18 nm, the control of etching is more critical. These thicknesses can absorb up to 20% variation in the etch rate.

For  $18 \text{ nm} < L_g < 90 \text{ nm}$ , the following film thicknesses proved to be effective choices:

$$K = 90 \text{ nm}$$

$$L_g \approx K$$

$$Z_{sp} \approx L_g$$

$$Z_r \approx 1.8 \times K$$

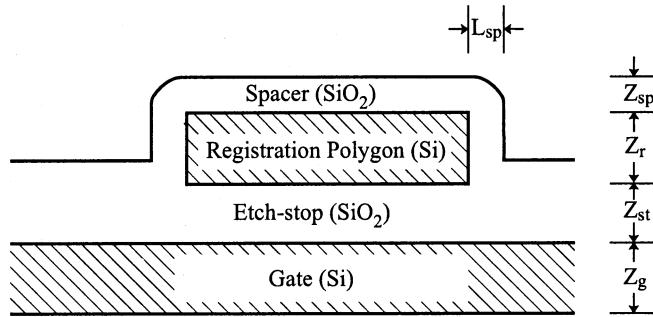
$$Z_{st} \approx 1.4 \times K$$

We stop the silicon etching depending on the automatic detection of etching completion, Figs. 5.3, 5.4. Typically, silicon etching tools feature an ‘end-point’ detection capability feature that reports etching completion by monitoring the spectrum of the etching byproducts to determine the composition of the etched film. Changes in the spectral intensity indicate etching completion. Silicon dioxide etching is estimated from a pre-calculated etching duration, based on a calibrated etching rate.

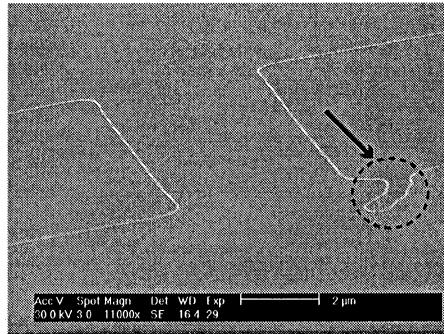
### 5.3.10 Defect tolerance

Edge-defined features surround the registration polygon. When a defect is present in creation of the registration polygon, the edge-defined features still conform to the registered edge.

Figure 5.17 shows an irregular defect in an edge-defined structure. This defect is believed to have resulted from particle contamination introduced during the

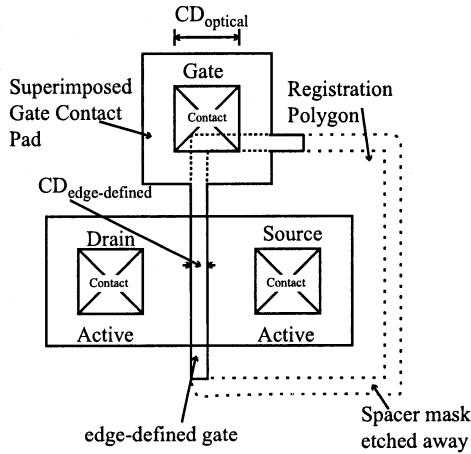


**Figure 5.16:** Defining film-stack thicknesses for an edge-defined process.



**Figure 5.17:** Defect tolerance of the edge-definition process. scale=2 $\mu$ m. The encircled kink is the result of a particle defect in the registration lithography process. The edge-defined structure is still electrically continuous.

registration photolithography process. The edge-defined structure conformed to the defect, however. In the case of an edge-defined structure formation, this behavior results in a radially displaced but electrically continuous structure. This is particularly important for the application of edge-defined processing to pattern transistor gates, where electrical continuity is critical. It also suggests an adaptation of the edge process to enhance active device channel performance, described in Section 5.4.3.



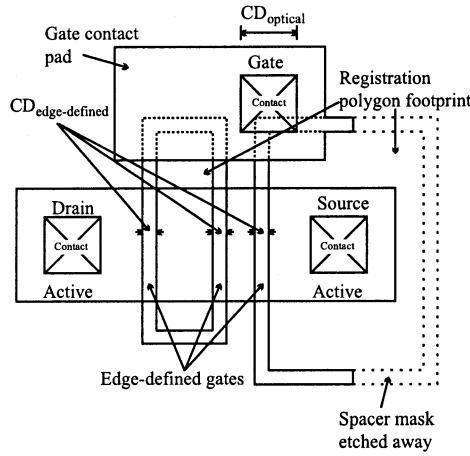
**Figure 5.18:** A single edge-defined gate finger. A large portion of the edge-defined ring can be etched away.

## 5.4 Mask Layout Techniques for Edge-Defined Lithography

The edge-defined patterning technique presented so far can be used for any general patterning application. It is particularly interesting, though, to pattern the gate structures of transistors using edge-defined processing. In a conventional transistor design, the gate is drawn directly during the mask layout. Use of edge-defined gates requires drawing the registration polygons instead. This drawing of registration polygons offers several interesting layout possibilities, which follow.

### 5.4.1 Footprint reduction

Since gates can now be made smaller than the features produced by simple photolithography, device sizes may be shrunk by moving the source and drain contacts closer to each other. Figure 5.18 shows a single-finger, edge-defined transistor. Note that the source-to-drain contact separation can be reduced for area savings.



**Figure 5.19:** A triple-front-gate device with three edge-defined transistor fingers. A single registration polygon defines two gate fingers; a second registration polygon is used for the third finger.

### 5.4.2 Multiple gates

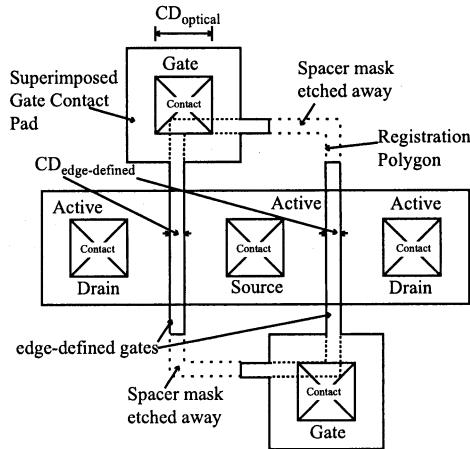
An edge-defined process inherently results in ring-like structures around the registration polygons. In order to make multiple isolated gates from a single edge-ring, another mask level is needed to remove the edge-ring when necessary.

While in some cases we must remove portions of the edge-ring, it is also possible to make use of a single edge-ring to define multiple gates, as shown in Figs. 5.19 and 5.20. Figure 5.19 shows the layout of a transistor with three edge-defined fingers. Note that two registration polygons are used. A single gate pad connects the three gates together. The second edge-ring is cut to create the third transistor gate.

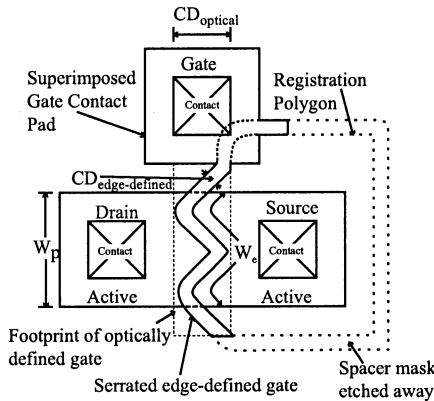
Figure 5.20 shows a layout for creating multiple devices using just one edge-ring. A single registration polygon defines the edge ring. Two separate gate pads are superimposed. The edge-ring is then cut into two transistors.

### 5.4.3 Serpentine edges

Edge-defined techniques present another interesting possibility. The edge-defined gates can be intentionally widened within a small footprint, to enhance the drive current of a transistor.



**Figure 5.20:** Two edge-defined transistors fabricated with a single registration polygon.



**Figure 5.21:** An edge-defined serpentine gate structure for drive-current enhancement. Cf. Fig. 5.17.

For example, Fig. 5.21 shows a serpentine gate structure. This serpentine structure can be produced by defining the outline of the serpentine structure arbitrarily, so long as it lies within the registration polygon.

This type of gate can be formed with an elongated gate width within a small footprint, which is a desirable feature. This geometry comes about because the edge-defined features can be smaller than the optically-defined features without etching optically-patterned photoresist. In Fig. 5.21, the serpentine gate structure stays within the footprint of an otherwise optically-defined gate. The effective width of the gate is  $W_e$ , which is longer than its physical width  $W_p$ . The contact holes are

defined optically; therefore, the minimum achievable critical dimension is based on the resolution of the optical system,  $CD_{\text{optical}}$ . In 130 nm and smaller CMOS technologies, the minimum dimension is achieved after overetching optically patterned photoresist.

#### 5.4.4 Experimental limitations

This work has been carried out at the Stanford Nanofabrication Facility (SNF). The shared nature of the facility results in frequent process variations. Etching and deposition rates vary from time to time, because of equipment usage by other projects, and because of cleaning schedules.

*With the necessary pieces now in place, we next show how practical, potentially manufacturable ULP circuits can be fabricated with critical feature sizes so small that until recently they would have been considered unattainable.*

## Chapter 6

### Device and Process Design

Theory and general observations are all very fine, but they affect the state-of-the-art and, beyond that, they affect human welfare only to the degree that they can be applied to yield practical results. Thus, with an understanding of edge-defined nanolithography processing and of the properties of the film surfaces, we now apply this understanding to the design and fabrication of devices having desirable features for ultra-low-power (ULP) technology. Implementing ULP technology leads to devices that possess threshold voltage ( $V_T$ ) adjustability in response to bias changes on the back-gate. In addition, we want these devices to be compatible with 3-D device stacking, and to have gate lengths ( $L_g$ ) shorter than 100 nm.

Toward the goals above, we note the following processing restrictions. Short channels are achieved using edge-defined lithography as described in Chapter 5. Compatibility with 3-D processing requires a transistor structure that can be vertically stacked in layers, and also a restricted thermal budget during fabrication to avoid long exposure to temperatures higher than 600 °C. Thin-film transistors (TFTs), especially those with ultra-thin bodies (UTB), are the most suitable structures for vertical stacking as well as for delivering optimal performance.

Conventional TFTs which are typically designed as three-terminal devices, do not, by design, possess a threshold-voltage dependence on the applied back-gate bias. In this chapter we present the design and fabrication of a TFT device with an insulated back-gate. The back-gate is used to adjust the  $V_T$  of these devices. The front-gates

are short-channel edge-defined features. These devices serve as prototypes to test the viability of the individual technology modules developed previously.

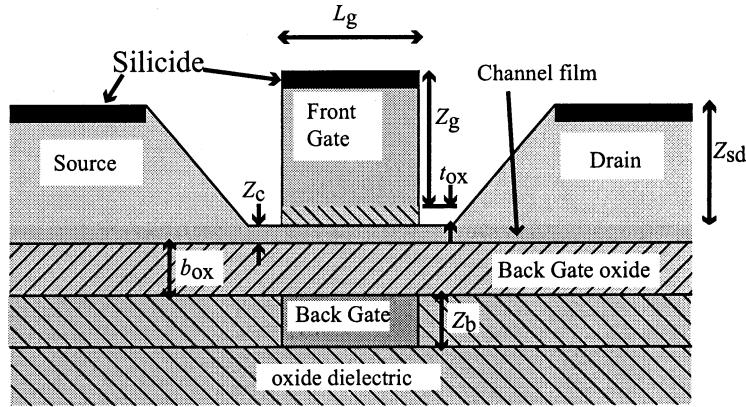
We focus, in particular, on the question of the design of the device structure and the processing flow. To begin, we introduce an ideal, and a device structure that addresses the issues at hand. Then we identify the manufacturing challenges of fabricating such a device. Subsequently we discuss device performance parameters, especially from the system-power-dissipation point of view. A sensitivity analysis is then carried out for the device. Finally the processing flow for the device fabrication is outlined.

## 6.1 ULP Device Structure

### 6.1.1 An ideal ULP device

An ideal implementation of a tunable, 3-D-compatible TFT is shown in Fig. 6.1. This device is similar to a conventional TFT in many ways. It differs from a conventional TFT in that an insulated back-gate electrode is provided for adjusting the on/off behavior. The front-gate and the back-gate have dielectrics of different thicknesses between each and the channel. This device incorporates a raised source and drain structure in order to reduce series resistance; the raised source and drain structures are tapered in order to reduce gate-to-source and gate-to-drain parasitic capacitances. The thin silicon (Si) channel offers the benefits of a fully-depleted SOI structure, including very little parasitic capacitance, complete device isolation, small leakage currents, and small subthreshold slopes. It also eliminates the mechanical stress-related issues due to commonly used trench-isolation techniques for transistors built in bulk Si.

The structural dimensions for a near-ideal device with 40-nm gate length are listed in Table 6.1. Fabrication of such a device requires development of a number of new techniques. In the following we outline the challenges that implementing this device design brings, and the possible solutions to each.



**Figure 6.1:** An ideal TFT device. An insulated back-gate electrode is used to adjust threshold and device variations.

**Table 6.1:** Optimal dimensions of a TFT with gate length=40 nm.

Gate length	$L_g$	40 nm
Gate height	$Z_g$	$\approx 100$ nm
Raised source and drain height	$Z_{sd}$	$\approx 100$ nm
Front-gate-oxide thickness	$t_{ox}$	$1 \sim 2$ nm
Channel Si thickness	$Z_c$	< 10 nm
Back-gate-oxide thickness	$b_{ox}$	$\approx 20$ nm
Back-gate Si thickness	$Z_b$	$\approx 50$ nm

### 6.1.2 Design challenges

The ideal device design shown in Fig. 6.1 presents a number of challenges. Several of these issues are specifically related to the engineering of devices with gate lengths  $< 50$  nm.

#### Channel film

Fabrication of smooth, single-crystal, ultra-thin Si films is a challenging task, as discussed in Chapter 4. Thin channel films can be obtained by using wafer bonding (Yang *et al.*, 1997), Si epitaxial lateral overgrowth (Denton *et al.*, 2001), and film deposition (Subramanian, 1998) processes.

Wafer bonding processing includes transfer of a thin single-crystal Si film from a sacrificial wafer to a second wafer, by first fusing the two together thermally, and then removing the sacrificial wafer in such a manner that it leaves a film attached to the other wafer surface. This process is complicated, and requires a separate set of sacrificial wafers for each stacked-device layer.

Selective epitaxial-growth processing can be used to grow single-crystal films vertically upwards from an underlying crystal seed. This process, however, requires temperatures above 700 °C, which are unsuitable for 3-D applications.

A third alternative is to use chemically vapor deposited amorphous Si (*a*-Si) films for the transistor channels, and subsequently apply a controlled crystallization technique such as metal-induced lateral crystallization, which can convert the deposited *a*-Si film to a crystalline film very effectively (Joshi & Saraswat, 2003) at relatively low temperatures (500–600 °C). With these techniques, the channel is deposited as *a*-Si film. A crystallization seed is either implanted or deposited. This step is followed by a low-temperature (500–600 °C) annealing process, which converts the film around the seed to a single-crystal form.

### Gate-dielectric films

Gate-dielectric films are typically 1–2 nm thick silicon dioxide (SiO<sub>2</sub>).<sup>1</sup> Unfortunately, for such small thicknesses, direct quantum-mechanical tunneling across the film produces leakage currents, and leads to poor insulating behavior of the gate dielectric. This problem is a subject of continuing research as part of studying high-k dielectric materials. It is beyond the scope of this dissertation, however.

For our device-fabrication experiment, we have avoided these issues related to the gate dielectric by using a ≈10-nm gate-dielectric film. The gate oxide is chemically vapor deposited (CVD), as opposed to thermally grown, because of the requirement to maintain a low temperature.

---

<sup>1</sup>Throughout this chapter, we will refer to silicon dioxide as just ‘oxide’ or SiO<sub>2</sub>.

### Raised source and drain

In order to decrease the drain-to-source series resistance, the drain and the source are flared outside the transistor region. Without such flaring, the series resistance could be sufficient to lower circuit speed. In addition, it is very difficult to fabricate a good metal contact with very thin films. Fabrication techniques such as selective germanium epitaxy can be used to fabricate raised source and drain structures.

### Implantation and annealing

Ion implantation at a low acceleration energy of  $\approx 1$  keV offers the benefit of precise dopant-profile engineering; low-energy implantation results in controlled-lateral ion ‘straggle’<sup>2</sup> during the implantation. Unfortunately, the lowest implantation energy available for this work is 13 keV. A screening film of  $\text{SiO}_2$  is used to reduce the effective energy of the implant, and implant-offset spacers are used to account for the lateral ‘straggle.’

### Contacts

Plasma-based etching is used to open contact holes in the dielectric film that covers the channel. An ideal contact-hole-etching process must be selective to the dielectric material and not etch the channel material, otherwise the contact etching may etch through the thin channel film. Therefore the channel film has to be thick enough in the source and drain region to withstand the contact-hole-etching process.

### Silicidation

Silicidation of the source, the drain, and the gate results in low values of series resistances. However silicidation has undesirable effects, such as inducing stresses and dissolving the entire Si film.

### Threshold-voltage adjustability

A primary objective of this work is to design and fabricate devices with threshold-voltage adjustability, where we define the threshold voltage as the gate voltage that

---

<sup>2</sup>Vertical implantation of the ions into the crystal lattice results in the lateral straggle of some of the implanted ions due to collisions with the lattice atoms.

results in channel charge inversion. When threshold-level voltage is applied to the gate, it opens a low-resistance path between the source and the drain.

For a conventional bulk CMOS transistor, threshold voltage is defined as

$$V_T = V_{FB} + V_S + 2 |\psi_B| + \frac{|Q_d|}{C_{ox}} \quad (6.1)$$

where  $V_{FB}$  is the flatband voltage and mainly depends on the work-function difference between the front-gate and the channel.  $V_S$  is the source voltage, and is typically zero. The variable  $\psi_B$  represents the relative potential of the carriers in the channel; for a doped semiconductor this is given by

$$\psi_B = -\frac{kT}{q} \left( \frac{N}{n_i} \right) \quad (6.2)$$

where  $N$  is the dopant concentration in the channel.  $Q_d$  is given by

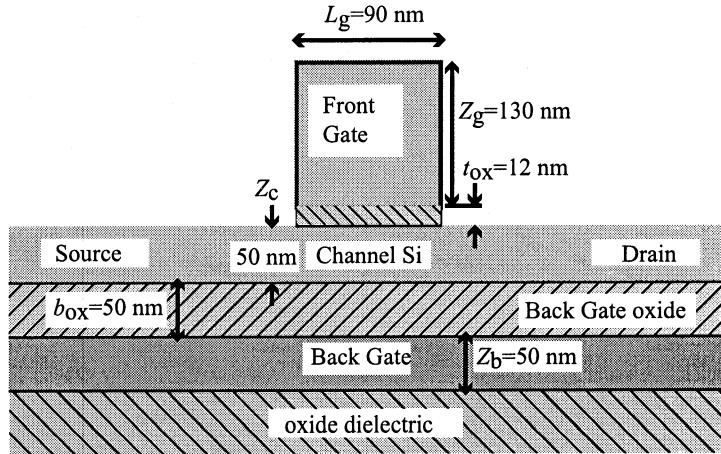
$$Q_d = -qN_a x_{dmax} = -\sqrt{2\epsilon_{sq} N (2|\psi_B| + |V_S - V_B|)} \quad (6.3)$$

These equations show that  $V_T$  can be adjusted by changing the work function, dopant concentration  $N$ , or back-gate bias ( $V_B$ ). Work function and dopant concentration methods of  $V_T$  adjustment can be used during device manufacturing. Back-gate bias, on the other hand, can be used to adjust  $V_T$  dynamically at run time. An ideal target for  $V_T$  is around 0 V, with a 100 mV/V adjustability in  $V_T$  from back-gate bias.

### 6.1.3 The prototype device

With the above mentioned issues in mind, we constructed a thin-film transistor that demonstrates our key ideas and contributions. This TFT is compatible with new technologies such as high-k gate dielectrics, raised source/drain, and metal-induced lateral crystallization processing.

Figure 6.2 shows a cross section of the prototype device fabricated in this work. Dimensions of this device are listed in Table 6.2. The back-gate electrode is a doped



**Figure 6.2:** TFT structure fabricated in this work, with  $L_g = 90$  nm. The channel film thickness is  $Z_c = 50$  nm. The back-gate electrode is insulated from the channel by a 50-nm thick back-gate-oxide film.

Si film with thickness  $Z_b = 50$  nm. The back-gate is insulated from the channel with a  $b_{\text{ox}} = 50$  nm back-gate-oxide film. The channel is a  $Z_c = 50$  nm thick *a*-Si channel. The channel film is thicker than optimal, but is able to support a good electrical contact. The front-gate oxide is a 12-nm LPCVD film. The front-gate length is  $L_g = 90$  nm, patterned with edge-defined lithography. The front-gate is  $Z_g = 130$  nm thick doped poly-Si. These values are summarized in Table 6.2.

The channel of this transistor consists of undoped Si. In a transistor with a doped short Si channel, a variation of even a ‘few’ doping atoms can cause noticeable threshold-voltage variations. An undoped channel does not suffer from

**Table 6.2:** Dimensions of prototype TFT with gate length=90 nm and a back gate.

Gate length	$L_g$	90 nm
Gate height	$Z_g$	130 nm
Front-gate-oxide thickness	$t_{\text{ox}}$	12 nm
Channel Si thickness	$Z_c$	< 50 nm
Back-gate-oxide thickness	$b_{\text{ox}}$	50 nm
Back-gate-Si thickness	$Z_b$	50 nm

such variations. Channel doping traditionally is used to control threshold voltages; but in the case of this transistor, these are controlled by back-gate bias.

## 6.2 Device Optimization

A 2-D commercial numerical simulator, Silvaco's 'Virtual Wafer Fab,' was used for device design and optimization. The processing flow as well as the device characteristics have been simulated. These simulations are based on an ideal device design in which the channel consists of single-crystal Si, a raised source/drain structure to reduce the series resistance, and a gate dielectric that is arbitrarily thin.

The simulation results are used to optimize the device design and to generate design rules for optimum performance. We evaluated a comprehensive set of simulations for NNP, NNN, PPN, and PPP structures.<sup>3</sup> These different combinations were investigated to understand any device sensitivities to the work function of the front-gate or back-gate. Threshold voltage, subthreshold slope, on-current, and off-current are extracted for all four types of structures for different assigned values of  $t_{\text{ox}}$ ,  $Z_c$ , and  $b_{\text{ox}}$ .

### 6.2.1 The device design rules

Based on the numerical simulations, we developed device design rules for the ideal device shown in Fig. 6.1. These design rules dictate the thicknesses of front-gate oxide ( $t_{\text{ox}}$ ), channel ( $Z_c$ ), and back-gate oxide ( $b_{\text{ox}}$ ) for a given size of gate length ( $L_g$ ).

For a given device gate length,  $L_g$ ,

$$t_{\text{ox}} = \frac{1}{20}L_g \longleftrightarrow \frac{1}{10}L_g,$$

$$Z_c = 3 \times t_{\text{ox}} \longleftrightarrow 4 \times t_{\text{ox}},$$

$$b_{\text{ox}} = 5 \times t_{\text{ox}} \longleftrightarrow 10 \times t_{\text{ox}}.$$

---

<sup>3</sup>NNP, e.g., means that the gate is n-type, the source and the drain are n-type, and the back-gate is p-type. First letter describes the doping type of the front-gate, second letter describes the doping of the source and the drain, and the third letter describes the doping of the back-gate.

**Table 6.3:** Device dimensions used for sensitivity analysis.

Gate length	$L_g$	40 nm
Gate height	$Z_g$	$= 1.4 \times L_g \approx 56 \text{ nm}$
Front-gate-oxide thickness	$t_{\text{ox}}$	1.5, 2, 3 nm
Channel Si thickness	$Z_c$	6, 8, 10 nm
Back-gate-oxide thickness	$b_{\text{ox}}$	17.5, 20, 22.5, 25 nm

Back-gate Si thickness ( $Z_b$ ) is optimized for a low-resistance path to the back-gate terminal. The gate-film thickness is based on the design rules used for edge-defined patterning.

These results are very intuitive: the front-gate oxide thickness should be  $\frac{1}{10}$  to  $\frac{1}{20}$  the gate length. A thin gate dielectric is required for good front-gate control over the channel. The channel film should be about three or four times thicker than the gate dielectric.

The thickness of the back-gate oxide is designed for optimum control of the threshold voltage. A threshold-voltage adjustability of 100 mV per volt of back-gate bias is optimum for ULP applications. If the adjustability is smaller than the optimum value then undesirably high values of back-gate bias are needed. On the other hand, if the adjustability is larger than the optimum value then noise starts to affect threshold-voltage.

### 6.2.2 $L_g = 40\text{-nm}$ device design sensitivity

Here we present the device design sensitivity analysis used for 40-nm gate length NNP structures based on device simulations.

The structural dimensions for the devices in the sensitivity analysis are listed in Table 6.3.

Figure 6.3 shows threshold voltages based on simulated  $I_d$ - $V_g$  device curves. The threshold voltage is the voltage-axis intercept along the maximum slope line of the

$I_d$ - $V_g$  curve. Threshold voltages are simulated for a device in the linear region,  $V_D = 0.1 V$ , as well as in the saturated region,  $V_D = 1.0 V$ .

Figures 6.3, 6.4, 6.5, and 6.6 show various values of extracted device parameters vs. thickness of front-gate oxide ( $t_{ox}$ ), thickness of channel ( $Z_c$ ), and thickness of back-gate oxide ( $b_{ox}$ ). The first of these, Fig. 6.3, shows the threshold-voltage variation. Threshold voltages are determined with and without 1 V of back-gate bias ( $V_b$ ), for  $V_D = 0.1$  and  $1.0 V$ .<sup>4</sup>

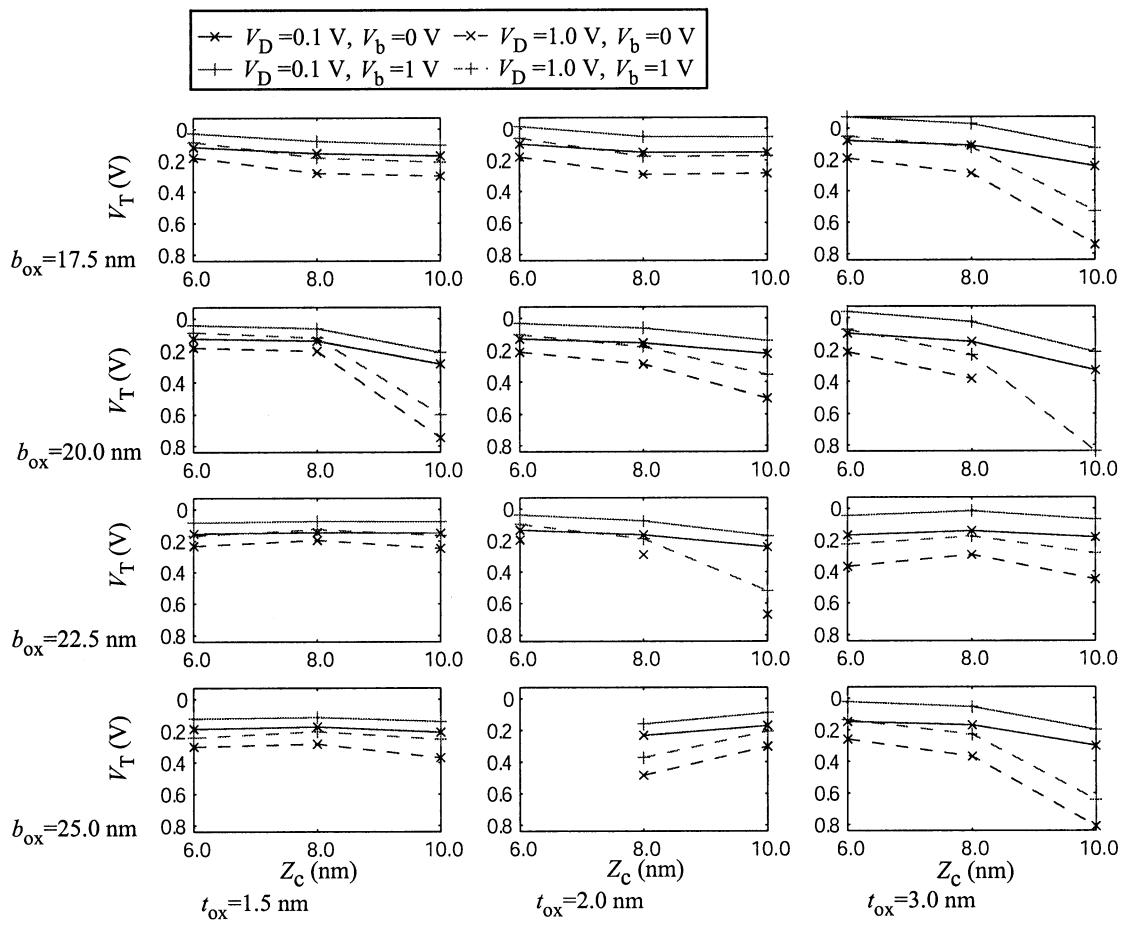
The  $V_T$  sensitivity to the device dimensions is clear from Fig. 6.3: thinner channels result in lower  $V_T$ , and thicker front-gate oxide films result in higher threshold voltages as expected. We also note that the use of thicker front-gate oxide results in better channel control by the back-gate bias. This is evident from the observation that for thicker  $t_{ox}$ , the same 1 V of back-gate bias results in higher shift for  $V_T$ .

Figure 6.4 shows subthreshold slope variation with device dimensions. Subthreshold slope is calculated without and with 1 V of back-bias ( $V_B$ ), and plotted for  $V_D = 0.1 V$ , i.e., the linear region, and  $V_D = 1.0 V$ , i.e., the saturated region. The target for subthreshold voltage is 60 mV/dec. Subthreshold behavior is most sensitive to the channel thickness and front-gate oxide thickness. Thinner channels and thinner gate dielectric favor lower subthreshold slope.

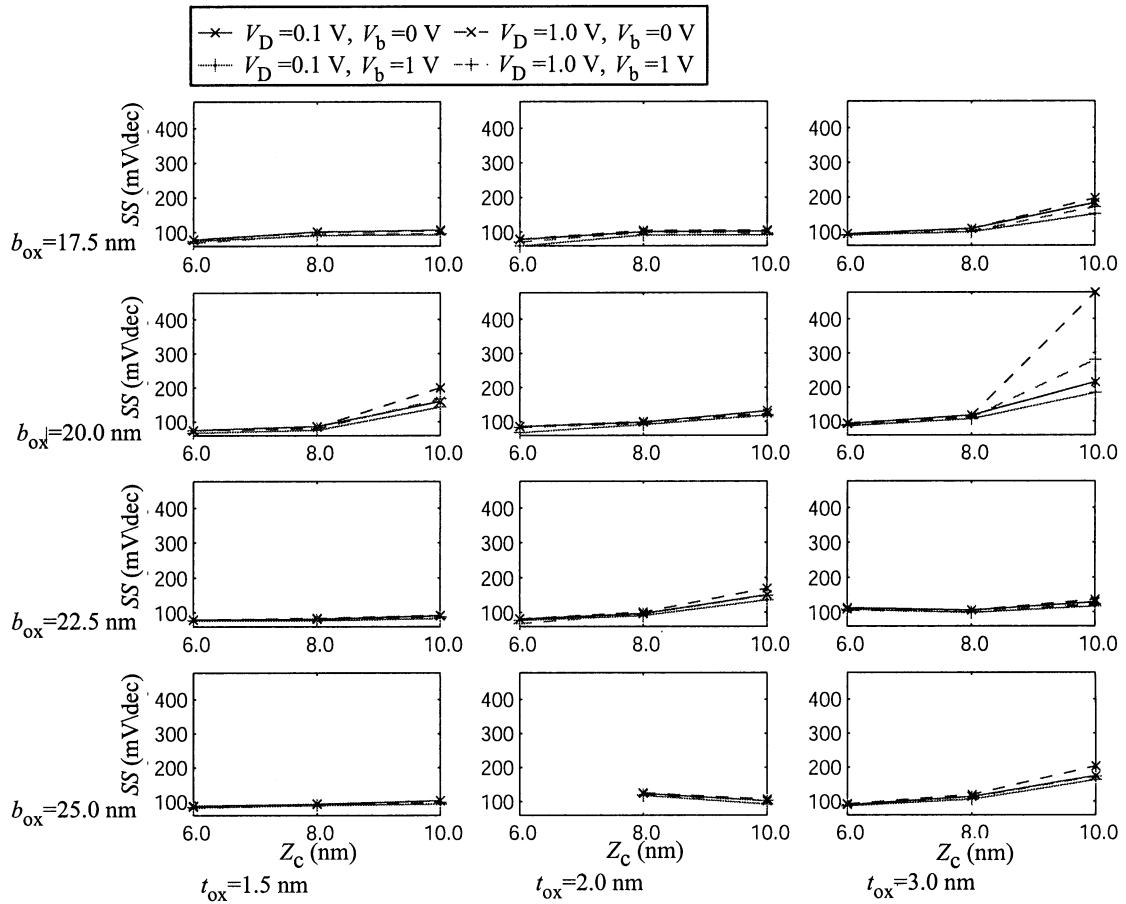
Figure 6.5 shows the on-current variation with device dimensions. For  $t_{ox} = 1.5 \text{ nm}$ , the values of on-current are relatively higher than those for thicker  $t_{ox}$ , which is intuitive; thinner front-gate dielectric results in lower inversion charge density, and hence lower current. Figure 6.5 also shows that the value of on-current decreases with back-gate bias. Note that back-gate oxide thickness  $b_{ox}$  does not effect the on-current significantly, since the on-current is mainly determined by the front-gate and

---

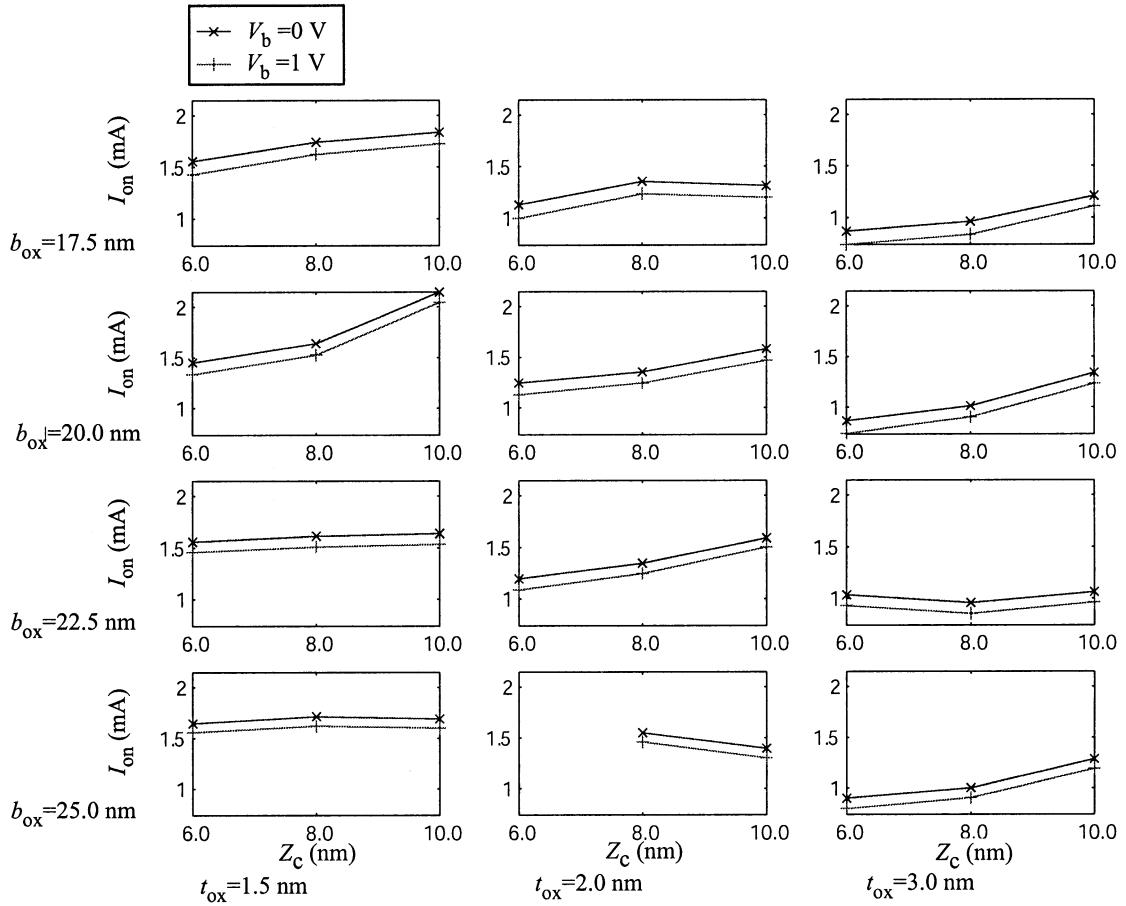
<sup>4</sup>Note that the absolute values of  $V_T$  for  $V_D = 0.1 V$  are not significant; we extract these for the sake of completeness. There are no dopants in the channel, therefore the threshold voltage is based only on the work functions as well as fixed charges. With this simulation we only analyze the device sensitivity to its geometrical dimensions, therefore we do not simulate any fixed charges. Absence of fixed charges is evident from the negative values of  $V_T$ . The absolute values of threshold voltages in our simulations, therefore, are not accurate. It is desirable to achieve transistors with static  $V_T$  in the range of 0–0.2V for ULP applications. The device work-function engineering and interface engineering are the main factors in achieving such low threshold voltages.



**Figure 6.3:**  $V_T$  variations vs. device dimension, for a 40-nm NNP structure. This  $V_T$  sensitivity analysis is based on combined processing and device electrical behavior simulations. See Table 6.3 for parameter definitions. Data points for  $Z_c = 6.0$  nm,  $t_{ox} = 2.0$  nm, and  $b_{ox} = 25.0$  nm, were not generated in the simulation.



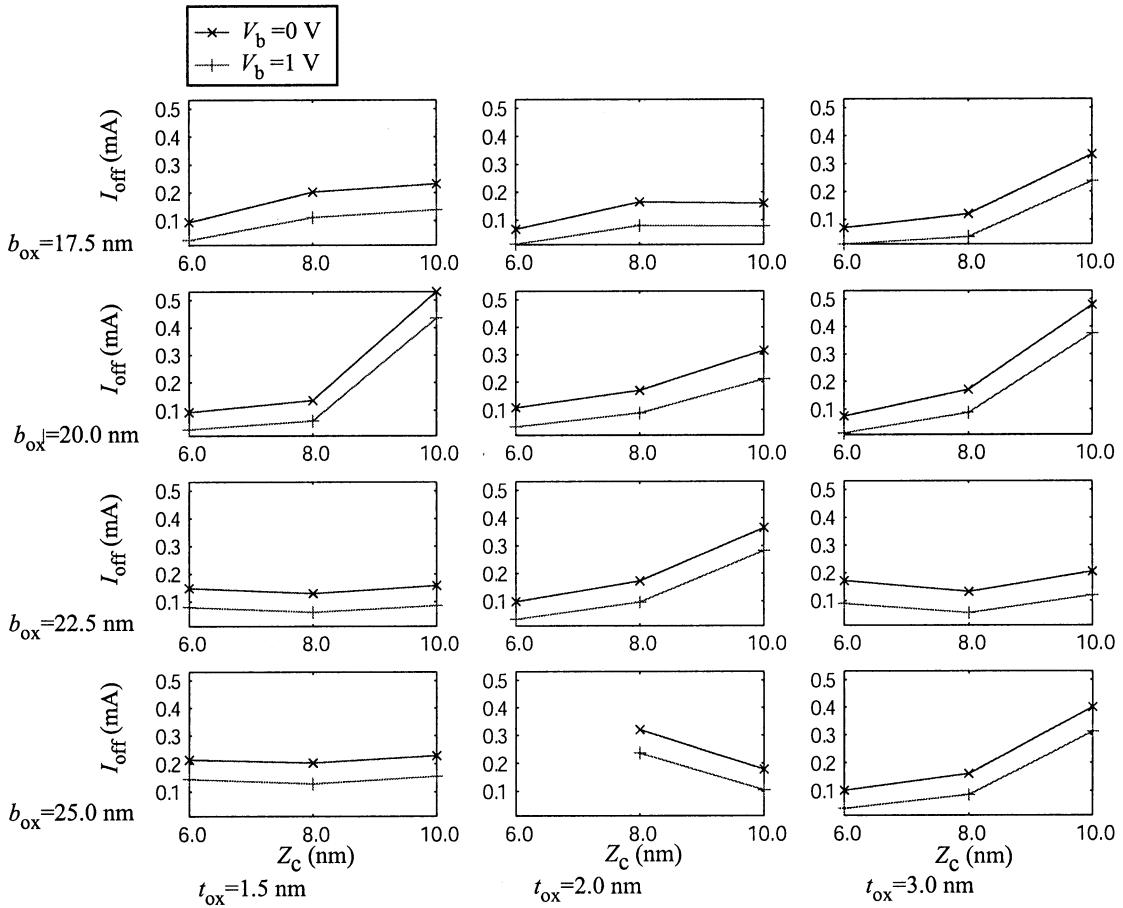
**Figure 6.4:** Subthreshold slope ( $SS$ ) variation vs. device dimension for a 40-nm NNP structure. Data points for  $Z_c = 6.0 \text{ nm}$ ,  $t_{\text{ox}} = 2.0 \text{ nm}$ , and  $b_{\text{ox}} = 25.0 \text{ nm}$ , were not generated in the simulation.



**Figure 6.5:** On-current ( $I_{on}$ ) variation vs. device dimension for a 50-nm NNP structure. Data points for  $Z_c = 6.0$  nm,  $t_{ox} = 2.0$  nm, and  $b_{ox} = 25.0$  nm, were not generated in the simulation.

front-gate-dielectric thicknesses. Interestingly, the on-current increases for thicker channels,  $Z_c > 6.0$  nm, due to the lower channel resistivity. This result should be understood, however, in the context of the subthreshold slope that degrades with the thicker channel films. The decrease in the subthreshold slope is significant enough for thicker channels to make these unsuitable for applications that require minimization of subthreshold leakage currents.

Figure 6.6 shows the off-current sensitivity with variation in device dimensions. The figure shows the extracted value of the off current without and with back-gate bias for different values of  $t_{ox}$ ,  $b_{ox}$ , and  $Z_c$ . Note that variations in the front-gate dielectric thickness,  $t_{ox}$ , and back-gate dielectric thickness,  $b_{ox}$ , do not greatly affect



**Figure 6.6:** Off-current ( $I_{\text{off}}$ ) variation vs. device dimension for a 50-nm NNP structure. Data points for  $Z_c = 6.0 \text{ nm}$ ,  $t_{\text{ox}} = 2.0 \text{ nm}$ , and  $b_{\text{ox}} = 25.0 \text{ nm}$ , were not generated in the simulation.

the off-current. Off-current, however, increases for thicker channels,  $Z_c > 6.0 \text{ nm}$ , mainly due to the current paths being farther away from the front-gate in the case of thicker channels. This result is consistent with the subthreshold slope sensitivity degrading with the thicker-channel films.

### 6.3 Device Processing Flow

Figure 6.7 shows key processing steps for device manufacturing. These steps represent a device processing flow comparable to a standard manufacturing flow, but with certain exceptions and key features. We focus on the distinctive features of this new

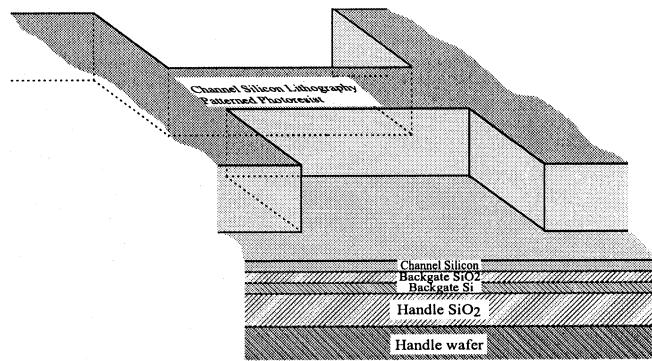
processing flow. Figure 6.7(a) shows the initial film stack. Here the Si wafer is used only as a handling mechanism; therefore it is referred to as a ‘handle wafer.’ The handle wafer is subjected to thermal oxidation at 850 °C to grow a 1000-nm thick SiO<sub>2</sub> film. This film is also used to support and handle the devices—hence it is referred to as the ‘handle oxide.’ Next we deposit the back-gate electrode Si film on the handle oxide. This film is degenerately doped in areas where the transistors are located. The back-gate electrode is shared among a group of transistors that belong to the same signal or power domain. A 50-nm thick *a*-Si film, ion implanted with boron and thermally crystallized, is used as the back-gate electrode. In our device, we estimate the mean thickness of the back-gate electrode to be around 43 nm using reflectance-based thickness-measurement equipment. Alternatively the crystallization step can be performed at the end of the processing flow.

The next step is formation of the back-gate oxide, SiO<sub>2</sub>, which is deposited with LPCVD as a 50-nm thick film. The channel-to-be is a deposited Si film on top, using an LPCVD process at 550 °C. The channel film is 50-nm thick, and is deposited in an amorphous state. Figure 6.7(a) also shows the channel lithography step with patterned photoresist located on top of the initial film stack. The pattern defines the active areas.

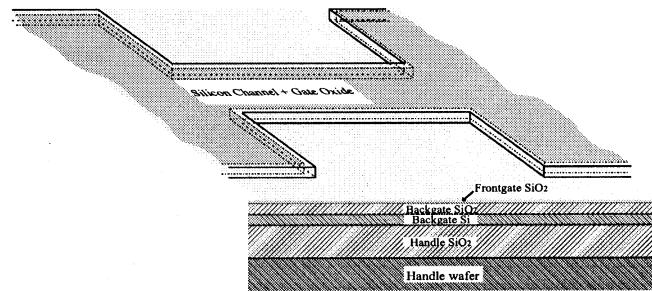
Figure 6.7(b) shows the patterned channel after gate-oxide growth above the channel. In this process the gate oxide is 12-nm thick LPCVD SiO<sub>2</sub> film, deposited at 450 °C.

The edge-defined gate patterning process of Chapter 5 is then used to form the transistor gate, as seen in Figure 6.7(c). The gate contact is placed on a gate pad. The gate pad is superimposed on the edge-defined gate, and is etched in the same Si film that is used for the edge-defined gate. Therefore both the gate pad and the gate are electrically connected. The gate-contact metal is placed on the gate pad. The source- and drain-contact metals are placed on the channel film. The back-gate contact metal is placed next to every transistor.

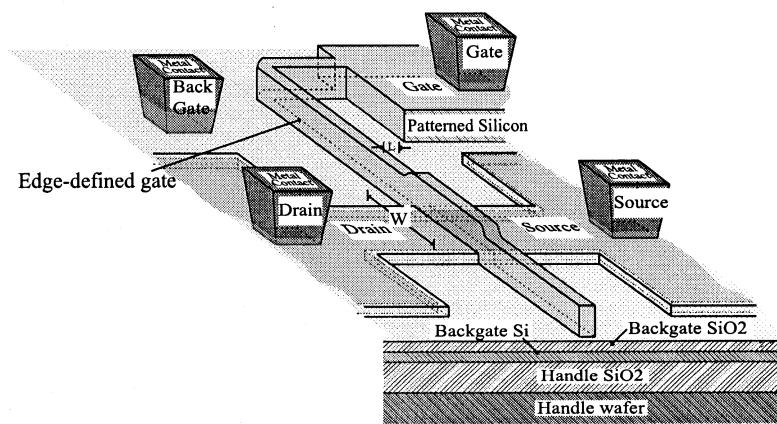
With this overview, we now describe important processing features in detail.



(a) Edge-defined 180nm line-width. The spacer mask is visible on top of the pattern.



(b) Edge-defined rings around the registration polygon phantom.



(c) The finished TFT with an edge-defined gate and metal contacts.

**Figure 6.7:** Edge-defined transistor fabrication. See beginning of Section 6.3 for explanation.

### 6.3.1 Process simulation

Figure 6.8 shows results from a numerical simulation of the processing flow for the prototype device. Insets 1–18, described below, show the edge-defined patterning process in the context of transistor formation. Features depicted are to scale.

Inset–1 shows the cross section of the wafer with back-gate Si film, followed by back-gate oxide film and channel *a*-Si film.

Inset–2 shows the gate-dielectric film added.

Inset–3 shows the gate-Si film added.

Inset–4 shows the stop-oxide film added.

Inset–5 shows the film stack after registration-Si deposition.

Inset–6 presents a block of patterned photoresist for sidewall registration.

Inset–7 presents the structure after the registration-Si film is etched, in order to transfer the registration pattern over to the registration layer, and the photoresist block has been removed.

Inset–8 shows the structure after spacer-oxide deposition.

Inset–9 illustrates the structures after spacer-oxide etching in order to form the sidewall spacer.

Inset–10 shows the sidewall-spacer mask after registration Si is stripped away.

Inset–11 illustrates the spacer further etched and translated down onto the gate Si.

Inset–12 shows the patterned gate formed using the spacer-oxide mask.

Inset–13 shows the deposition of the oxide film used as an ion-implantation-offset spacer.

Inset–14 shows the doping profile after the ion implantation step using arsenic at implantation energy of 13 keV, and a dose of  $2 \times 10^{15} \text{ cm}^{-2}$ . Note that this implantation step creates doped source/drain tips.

Inset–15 shows deposition of another oxide film, used as a second ion-implantation-offset spacer.

Inset–16 shows the dopant profile resulting from phosphorus-ion implantation at an implantation energy of 50 keV and a dose of  $3 \times 10^{15} \text{ cm}^{-2}$ .

Inset–17 shows the final device structure.

Inset–18 shows the simulated dopant profile after device annealing to crystallize the channel. This last step also redistributes dopants in the channel.

### 6.3.2 Processing details

#### Alignment scheme

The alignment targets are typically trench-type features a few microns wide, etched into the wafer at an early processing step. The trench outline is used for subsequent lithography-layer alignment. Edge-defined patterning processes require depositions and etching of three sacrificial films. These three films leave residues along topographical discontinuities. Any residue along the alignment target edges subsequently degrades alignment accuracy. We use multiple alignment targets in order to ameliorate this problem. The blank wafer is initially covered with a 1000-nm thick blanket of  $\text{SiO}_2$ , thermally grown in 850 °C-steam ambient. The initial alignment targets are etched 1000-nm deep into this  $\text{SiO}_2$  film. Secondary alignment targets are added at critical lithography levels. With the nominal-resolution 1  $\mu\text{m}$  Ultratech stepper available at SNF, we obtain an alignment accuracy of 200 nm.

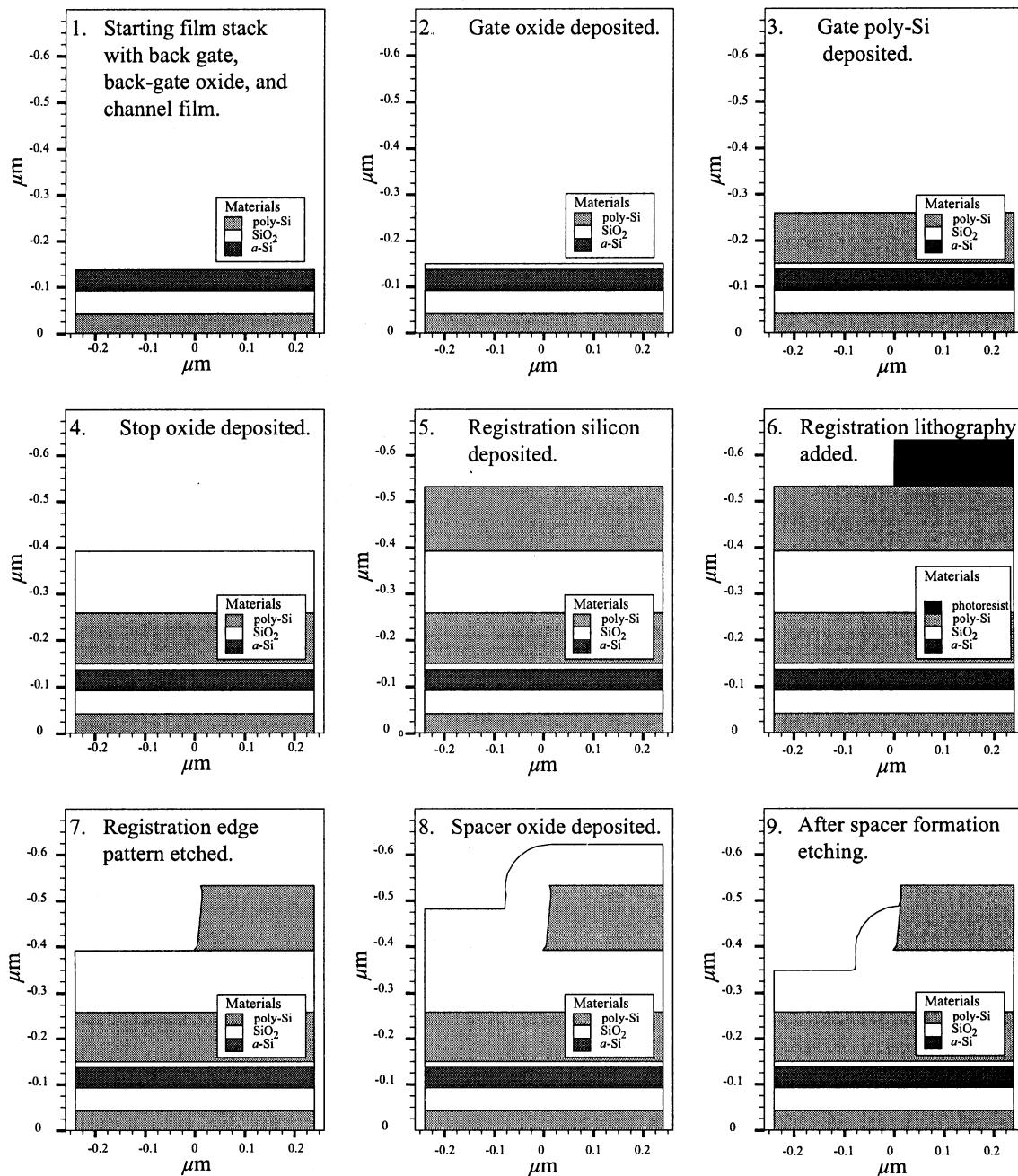
In order to measure the alignment accuracy, we have used a two-dimensional vernier-structure scheme, as shown in Fig. 6.9. The vertical and horizontal vernier structures are used to measure the lithography offset along both axes. The alignment offset measurement is used to control the photolithography process steps.

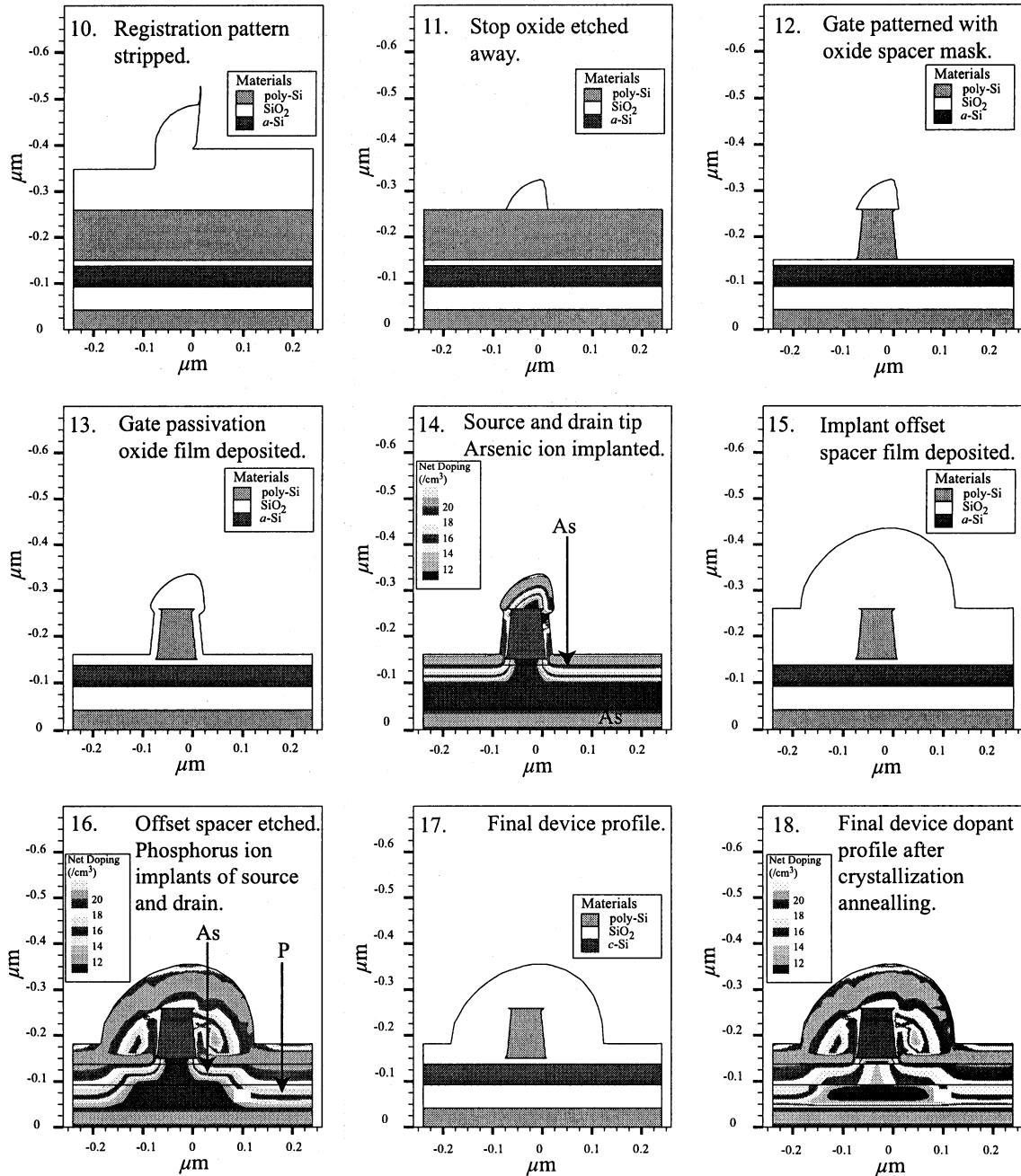
#### Channel formation

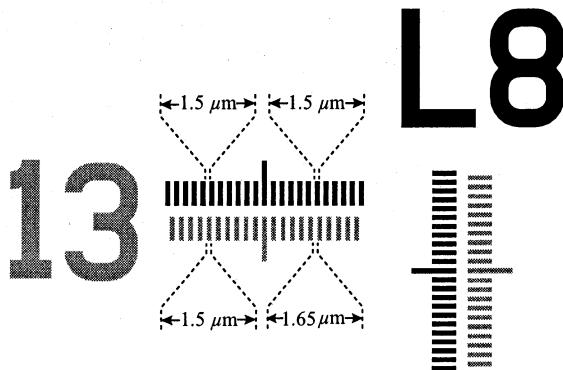
The transistor channel is an LPCVD deposited intrinsic *a*-Si film. Details of the channel film deposition are presented in Chapter 4. The completed channel film

*Next two pages.*

**Figure 6.8:** Simulated steps in edge-defined transistor fabrication. The simulated features shown were obtained with the *Silvaco Athena* 2-D process simulator.







**Figure 6.9:** Alignment vernier structures used to measure alignment. The vertical and horizontal verniers have the same spacing.

**Table 6.4:** Device dimensions used for sensitivity analysis.

Spacer film thickness	$Z_{sp}$	$= L_g = 90 \text{ nm}$
Registration film thickness	$Z_r$	$= 1.8 \times L_g = 160 \text{ nm}$
Etch-stop oxide film thickness	$Z_{st}$	$\approx 1.4 \times L_g = 120 \text{ nm}$

is a solid-phase crystallized poly-Si film. Finishing the crystallization of this film occurs near the end of the process. The channel film is initially deposited as *a*-Si, in anticipation of the final lateral crystallization processes (Subramanian, 1998).

### Gate patterning

The edge-defined gate-patterning process described in Chapter 5 is used to pattern the gate. Table 6.4 lists thicknesses of the films used for the edge-defined patterning process. Cf. Appendix B for processing details.

### Source and drain formation

The source and drain are formed by a dual ion-implantation scheme. For transistor features in the nanometer-scale range, the ideal energy for the first ion implantation is in the sub-keV range. Because we lack an ion-implantation setup that can operate at sub-keV energies, we used an ion-implantation step using arsenic ions at 13 keV with a dose of  $2 \times 10^{15} \text{ cm}^{-2}$  as the first step to define the source and drain extensions. We then deposited a 100-nm offset spacer film that serves as the source and drain implant

mask, helping to reduce the source- and drain-doping atoms from straggling off into the channel. A second ion-implantation step then follows, which uses phosphorus ions  $3 \times 10^{15} \text{ cm}^{-2}$  at 50 keV.

### Contact formation

Metal contacts are formed with a combination of dry-plasma and wet etching. Contact holes are etched through 250-nm thick  $\text{SiO}_2$ , the dielectric between the devices and the metal interconnect. Contact-hole etching must stop on the top side of the 50-nm thick channel Si film and 50-nm thick back-gate Si film. Plasma etching of  $\text{SiO}_2$  is used to etch contact holes almost, but not quite, all the way through the  $\text{SiO}_2$  layer. Contact-hole etching is completed with a diluted HF solution which etches away  $\text{SiO}_2$  faster than Si. A surfactant was added to the HF solution in order to improve the wetting angle of the solution. This addition also helps the solution wet the very small contact holes.

### Channel crystallization and activation

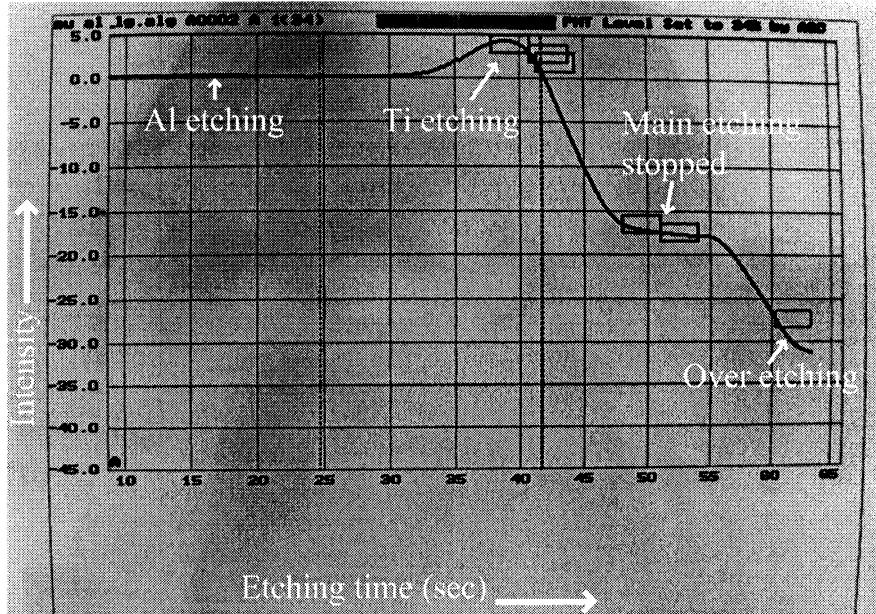
Amorphous-Si crystallization and dopant activation is achieved by rapid thermal annealing (RTA). RTA processing is designed to deliver a one-second flash RTA at 900 °C.<sup>5</sup> Unfortunately, calibrated RTA processing is not available; therefore, an educated estimate was used to fabricate the devices.

### Metal interconnect

The interconnection metal is deposited after completing the contact-hole formation and RTA crystallization. The interconnect system consists of a 50-nm titanium film topped by a 450-nm aluminum film, both deposited using sputtering processes. The titanium film serves as the ‘glue’ that adheres to the very thin channel Si film. After depositing a blanket of metal for the interconnect layer, the wafers undergo interconnect photolithography. Photoresist is used as the interconnect masking material during the metal etching process employed to pattern the interconnect. A

---

<sup>5</sup>Previously mentioned limitation of 600 °C only applies to longer duration temperature exposures. Usage of short-duration high-temperature RTA or flash annealing can be engineered to avoid significant dopant diffusion in the transistor channels.

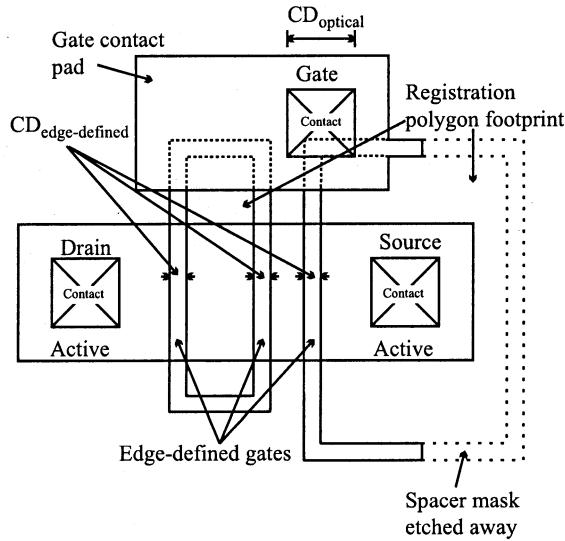


**Figure 6.10:** Metal etching. Image of the P5000 Applied Materials etcher console after completion of metal etching. The spectrometer monitors the intensity of light at  $\lambda = 400$  nm in the etching byproducts. Aluminum and titanium both emit light at about 400 nm; therefore, the sudden decrease of intensity indicates disappearance of Al and Ti, from which we infer completion of etching.

P5000 Applied Materials etcher is used for dry etching of the metal. Figure 6.10 shows an image from the optical emission spectrometer of the P5000, obtained while etching the metal interconnect. The spectrometer monitors the intensity of light at  $\lambda = 400$  nm in the etching plasma. Aluminum and titanium both emit light at 400 nm; the observation of a sudden decrease in intensity at 400 nm implies the disappearance of Al and Ti, indicating completion of the metal etch.

### Annealing

Final annealing of the completed device occurs at  $350^{\circ}\text{C}$  in a forming gas ambience, which provides hydrogen atoms that diffuse through the device structure. These diffusing H atoms terminate the dangling bonds at the Si–SiO<sub>2</sub> interface and reduce the fixed interface charge, which in turn lowers the manufactured value of  $V_{\text{T}}$ .



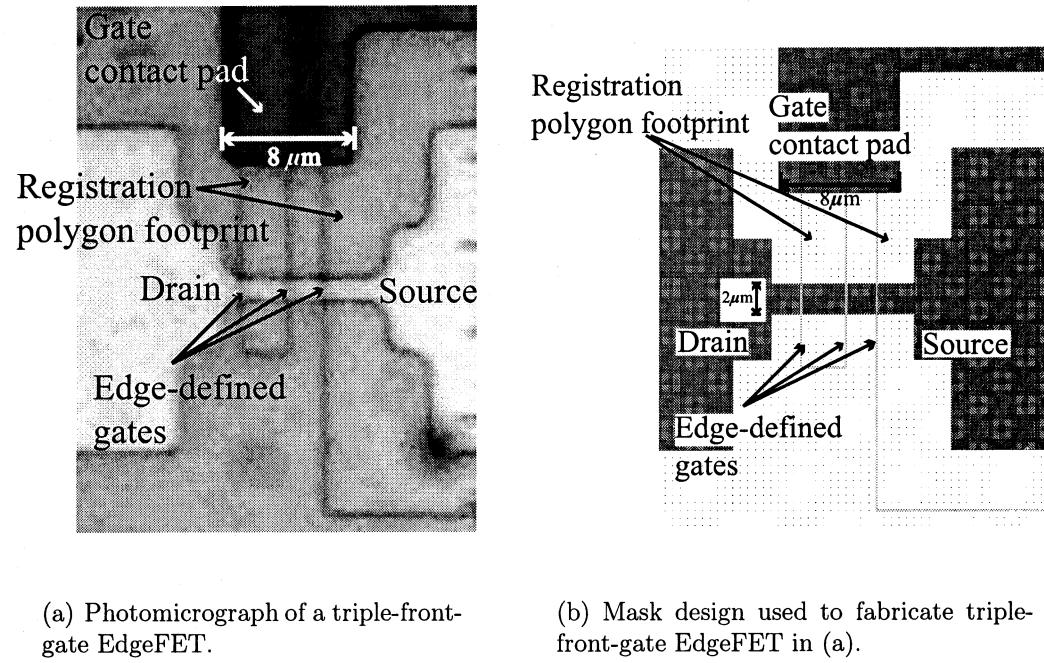
**Figure 6.11:** Drawn footprint of a triple-front-gate EdgeFET with three front-gate fingers. The critical dimension (CD) used as the transistor gate is defined using an edge-defined process, whereas the CDs for the metal contact holes, gate contact pad, and registration polygon are defined by photolithography. Two registration polygons are used to pattern three edge-defined front-gate fingers. The channel and the source and drain contact holes are also defined with photolithography.

## 6.4 The Completed Device

Our prototype devices are referred to as “edge-defined field effect transistors” (EdgeFETs). This section describes the physical structure of the  $L_g = 90$  nm prototype EdgeFET.

Figure 6.11 sketches the footprint of a triple-front-gate EdgeFET with series-connected transistor channels. The ‘active’ area, source and drain regions, and the contact holes are defined by photolithography. Two edge-defined gate rings that are constructed around two separate edge-registration polygons. The edge-defined spacer mask is removed selectively wherever it is not needed, as part of the fabrication processes. The edge-defined mask is merged with the gate-contact pad, which is defined by photolithography.

Figure 6.12 shows the footprint of a prototype triple-front-gate EdgeFET device with series-connected transistor channels. Figure 6.12(a) shows a photomicrograph of the actual device as built, whereas Fig. 6.12(b) shows the mask design used to build



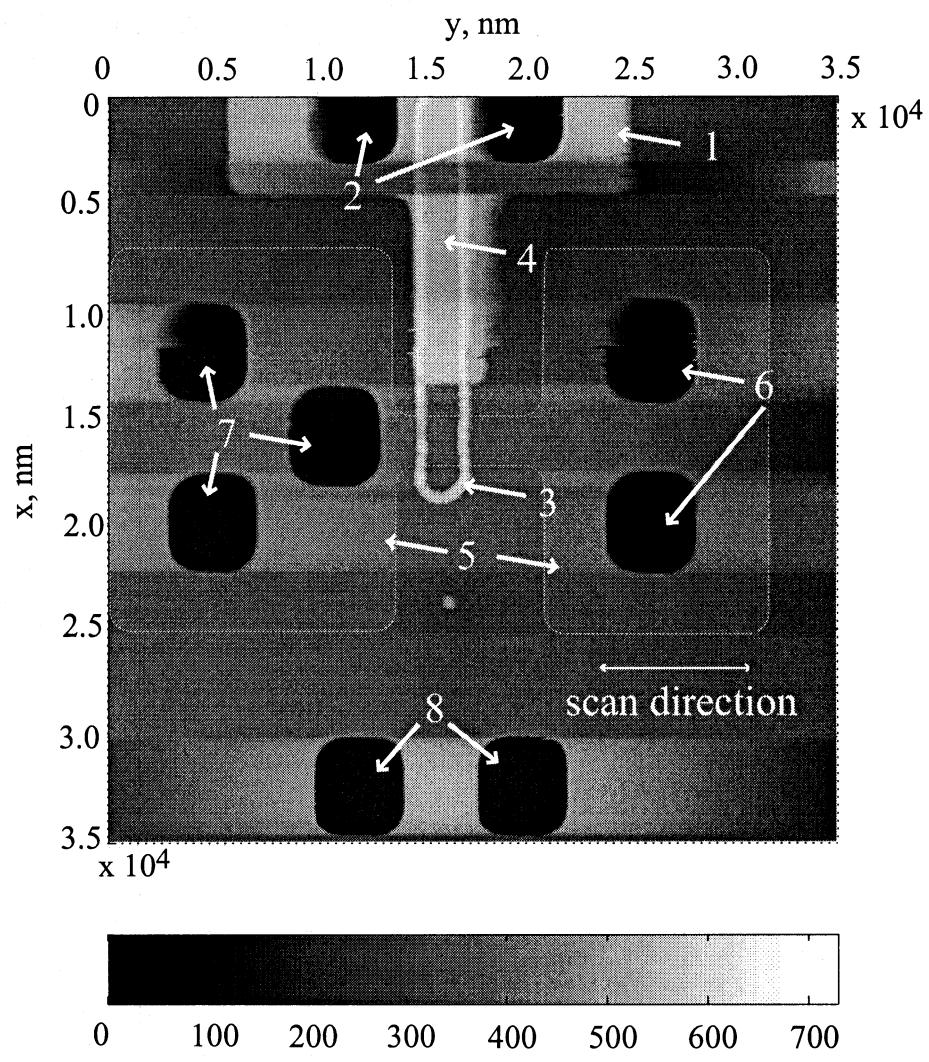
**Figure 6.12:** A prototype triple-front-gate EdgeFET with series-connected transistor channels.

this device. Again, this device was fabricated by combining edge-defined gates and photolithography-defined channels, contact holes, and interconnects. The active area and the contact holes are large photolithographic-size features. The two edge-defined gate rings are built around two separate edge-registration polygons. The edge-defined spacer mask is removed selectively where it is not needed; the edge-defined mask is merged with the gate-contact pad defined by photolithography.

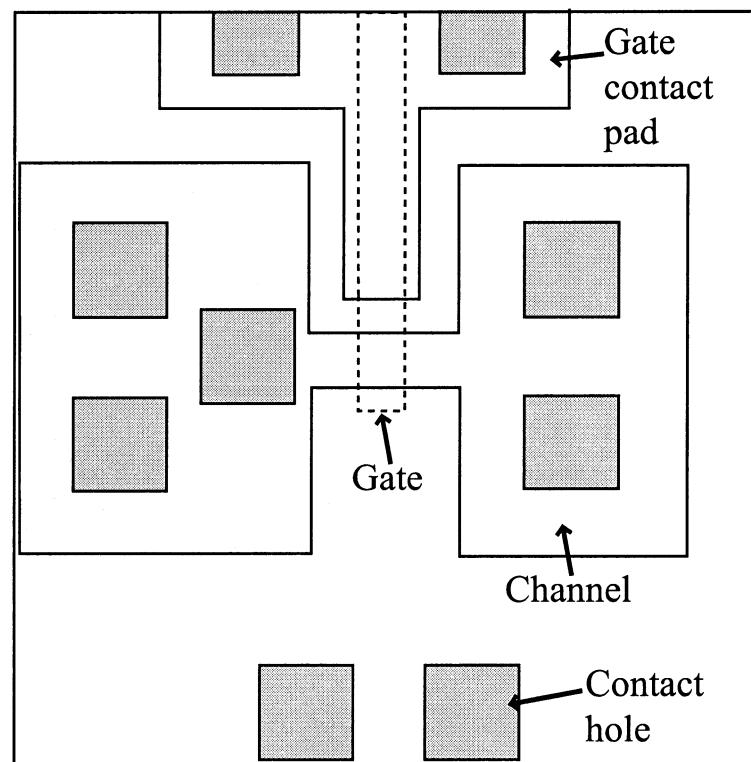
Figure 6.13(a) shows a top view of an atomic force microscope (AFM) scan of a prototype dual-front-gate EdgeFET. This scan shows the device immediately after etching the contact holes through the  $\text{SiO}_2$  dielectric layer, but prior to deposition of metal for electrical contacts. The gate contact pad is marked by label '1,' and contact holes to the gate are marked by '2.' The white ring, '3,' is an edge-defined gate ring; it results in two front-gate fingers across the channel, where it crosses the active area. The signature of the edge-defined gate is picked up by the AFM, because of its height due to the spacer mask. The edge-defined gate overlaps with the gate pad that is defined by photolithography in area '4.' The dumbbell-shaped structure,

*Next two pages.*

**Figure 6.13:** (a) AFM scan of a dual-front-gate EdgeFET after contact holes are etched through the  $\text{SiO}_2$  dielectric film, but before metal deposition. This device has two edge-defined front-gate fingers and the transistor channels are connected in series. 1–Gate contact pad. 2–Gate contact hole. 3–Edge-defined gate ring. The dimensions of the gate ring along the x and y axes are widened by the smoothing action of the finite size of the AFM tip. 4–Optically-defined gate. 5–Channel silicon; it is shaped like a dumbbell. The outline of the channel is digitally enhanced. 6–Channel contact hole. 7–Channel contact hole. 8–Back-gate contact hole. (b) The mask outline of the feature seen in the AFM scan.



(a)



(b)

‘5,’ is the active-area shape etched into the channel Si film. The channel film is only  $\approx$ 50-nm thick; therefore, the color difference with respect to the background is not very pronounced. The narrow portion of the dumbbell-shaped channel is the actual active area where the transistor is formed; the outer portions are only used for contacts. The channel contacts are marked by ‘6’ and ‘7.’ Contact holes appear dark because they represent deep trenching through the metal dielectric layer. Please note the vertically running bands of bands of discoloration. These bands are caused by uneven flattening of the overall AFM scan due to deep contact holes. Contact holes to the back-gate are marked by ‘8.’ Figure 6.13(b) shows the mask outline of the same feature scanned by the AFM.

*And now, ‘the proof is in the pudding.’ We turn from design-engineering issues to what for production ICs are product-engineering issues, of characterization of individual devices and of device-to-device comparisons.*

## Chapter 7

### Device Results

How well do these extremely small-geometry devices perform? How uniform are their characteristics across a die, and across a wafer? How can we best test them? What is the optimum roadmap to practical exploitation of this technological approach?

The devices fabricated according to the methods of Chapter 6 exhibit threshold-voltage adjustability by application of back-gate bias. The ability to adjust the device threshold voltage  $V_T$  by application of a bias control voltage,  $V_B$ , to the substrate beneath the control gate allows the VLSI systems designer to manage power consumption dynamically at run time, e.g., in response to changing demands by the overall operating system and application software. The primary tradeoff is in the balance between leakage power and switching energies in device and circuit operation (Svilan *et al.*, 2000). Such a capability to adjust  $V_T$  optimally is even more important for circuits stacked in 3-D, where heat removal is a pressing issue (Souri *et al.*, 2000). The devices described here have  $L_g = 90\text{--}360\text{ nm}$ , increasing in multiples of 90 nm. These prototype devices are of special interest in the contexts of 3-D silicon and of silicon-on-flexible-substrate. Similar techniques can be applied to fabrication of high-performance bulk-Si devices intended for use as the first layer in 3-D stacking. The device mobility and subthreshold behavior reported here are comparable to that reported in other work on poly-Si thin-film transistors (TFTs) (see e.g., Watts & Lee, 1993).

Our prototype devices are polycrystalline silicon (poly-Si) channel thin-film transistors with accessible back-gates, created in a 3-D-compatible IC process as described in Chapter 5. Gate patterning in these devices is achieved using a low-temperature-compatible edge-defined process. Edge-defined lithography is particularly advantageous for 3-D integration because it offers a cost-effective means to produce the multiple critical dimension (CD) mask layers required for stacked devices. In this patterning process a spacer is defined around a sacrificial registration block. The sacrificial registration block is then removed, leaving the outline of the registration in the form of a spacer. The spacer width is defined by the spacer-film thickness, which ultimately controls the final gate length,  $L_g$ , by acting as a final mask. A second pattern is superimposed upon the edge-defined spacer mask, thus combining fabrication of edge-defined and optically-defined features at the same level. The photoresist and spacer are used to pattern the final feature, which in this application is a control gate.

The TFT's active area comprises a 50-nm thick deposited amorphous silicon (*a*-Si) channel, which in a final step is crystallized by means of rapid thermal annealing. An underlying layer of boron-doped Si film serves as a global back-gate electrode; this layer is insulated from the channel by a back-gate oxide film of 50-nm thickness. A 12-nm blanket film of low-pressure chemically-vapor-deposited (LPCVD)  $\text{SiO}_2$  is added for the gate insulator after the active area is patterned. An *in-situ* phosphorus-doped *a*-Si gate film is deposited and patterned into 90 nm-length gates using edge-defined lithography. Arsenic ions are implanted to create the sources and the drains, whereas the gates are created as deposited *in-situ* doped amorphous Si. After contact holes have been created, the devices are subjected to 900 °C rapid thermal annealing (RTA) of a few seconds duration in order to activate dopants and to convert the channel *a*-Si into poly-Si form (cf. Page 116). The device processing does not include any dopant implantation for the channel, other than the unavoidable diffusion of arsenic ions from the source and drain ion implantation. The contact metal system consists of a 50-nm

**Table 7.1:** Sheet resistivity measurements after 1-sec RTA. Note that the metal interconnect processing is carried out after RTA.

Film Type	Film thickness	Sheet resistance kΩ/square
Metal	500 nm Al, 50 nm Ti	0.057 Ω
N-active	50 nm	5.33 kΩ
N+ Gate poly-Si	110 nm	95.7

thick titanium film overlain by a 450-nm aluminum film. After metal deposition and patterning, the devices are annealed at 350 °C in forming gas<sup>1</sup> for an hour.

We start with the results of sheet resistivity measurements of different films. Such measurements are important for the analysis of the data presented here. Next we discuss representative device curves measured for the fabricated device; we then present the effect of back-gate bias on the device behavior as measured for two different devices.

Section 7.4 presents the results of device performance across a range of operational device temperatures. Section 7.5 covers the intra-die device-performance variations and the explanation of that behavior.

## 7.1 Sheet Resistances

The sheet resistance,  $R_{\text{sheet}}$ , of the channel, gate, and metal films have been characterized using Kelvin-probe structures. Table 7.1 contains extracted values of sheet resistance for a wafer with 1-sec rapid thermal annealing.

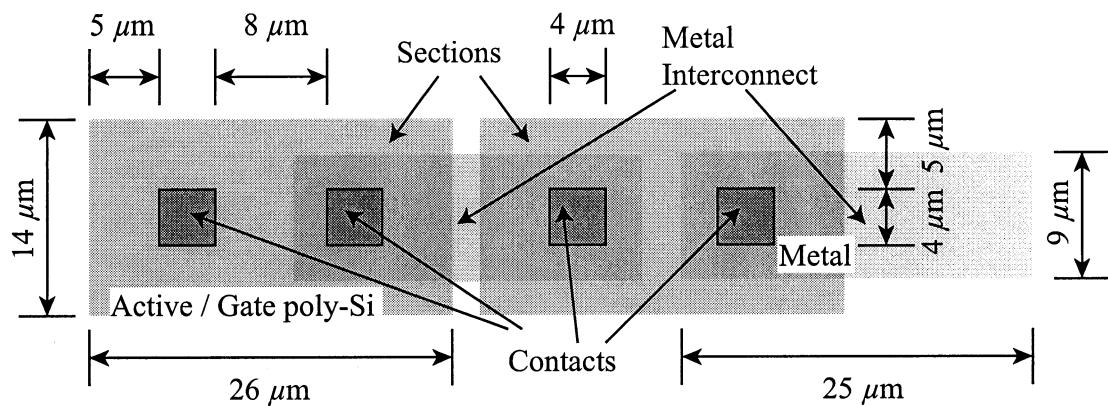
Table 7.2 contains values of sheet resistance for a wafer with 3-sec rapid thermal annealing. Notice that longer thermal exposure resulted in lowered sheet resistance, due to higher concentration of activated dopant atoms.

---

<sup>1</sup>Forming gas is a 5:1 mixture of N<sub>2</sub> and H<sub>2</sub>.

**Table 7.2:** Sheet resistivity measurements after 3-sec RTA.

Film Type	Film thickness	Sheet resistance kΩ/square
Metal	500 nm Al, 50 nm Ti	0.057 Ω
N-active	50 nm	4.93 kΩ
N+ Gate poly-Si	110 nm	86.4



**Figure 7.1:** Contact chain structure. The two sections of Si are connected serially through metal interconnects and contact holes. A serial chain of 500 sections is formed, and can be probed across 10, 20, 40, 100, and 500 sections.

Contact-chain structures have been fabricated to evaluate the integrity of contacts. Figure 7.1 illustrates the structure of the contact chain including unit sections. The chain consists of five hundred sections of Si, serially connected using metal interconnect. Each section is a rectangular piece,  $26 \times 14 \mu\text{m}^2$ , channel Si or gate poly-Si. These chains have probe points to measure the resistance of 10, 20, 40, 100, and 500 sections.

Table 7.3 gives the resistance of the test contact chains. These measurements are for a wafer subjected to 1-sec rapid thermal annealing. Resistance per section is calculated by dividing the total resistance measured across by the number of sections spanned. The resistance coefficient shows the change in the resistance value from the change in the voltage across, calculated as percentage change per volt.

**Table 7.3:** Contact-chain resistance.

No. of sections	Resistance/section kΩ	Resistance coefficient %change/V	Type
10	3.127	-0.307	N-active Si
20	3.648	-0.132	N-active Si
40	3.925	-0.061	N-active Si
100	4.060	-0.024	N-active Si
500	4.173	-0.005	N-active Si
500	0.079	-0.243	N+Gate poly-Si

## 7.2 Device Curves

The fabricated test die contains twenty-four prototype transistors with different channel widths,  $W_g$ , and channel lengths,  $L_g$ ; each wafer contains about 30 dice. Only eighteen transistors can be electrically probed, however due to a processing error. Table 7.4 shows the dimensions of test devices fabricated for prototyping purposes. Devices identified as “s6dvx” have a single edge-defined gate with  $L_g=90$  nm; devices identified as “s7dvx” have one or more edge-defined transistor channels in series. Channel lengths of these devices are multiples of 90 nm. The DC behavior of these devices is summarized by  $I_D-V_G$  and  $I_D-V_D$  curves which are measured directly for the devices on the wafers.

Figure 7.2 shows DC device curves of a representative single edge-defined gate with  $L_g=90$  nm and  $W_g=2000$  nm, at  $\approx 20$  °C. In Fig. 7.2(a), the family of  $I_D-V_D$  curves with solid lines represents the device behavior for zero back-gate bias. These curves do not show saturation of the drain current with increasing  $V_D$ , which indicates control of the channel conductivity by the drain voltage similar to that from gate voltage. This behavior may be the result of drain-induced barrier lowering (DIBL), as observed in standard CMOS technology, which is explained by sub-optimal device dopant profiles as a result of enhanced diffusion along grain boundaries. Uncontrolled inward diffusion of dopant atoms from the source and the drain results in shortening

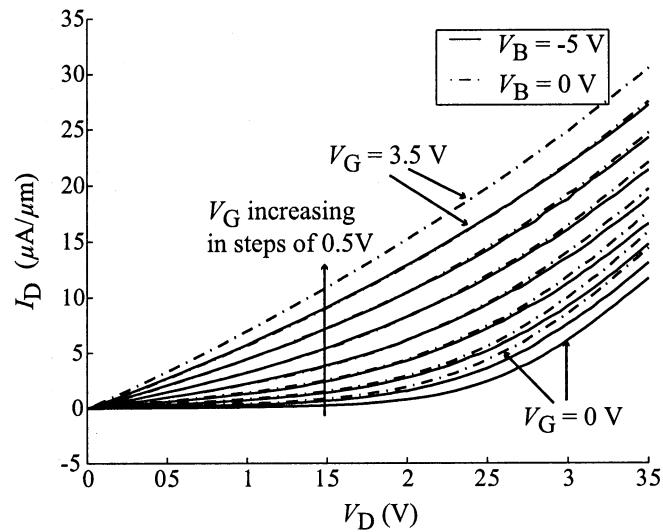
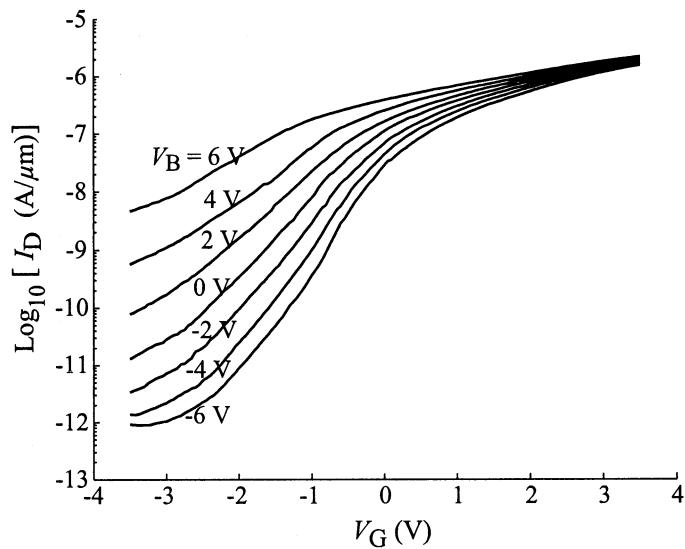
**Table 7.4:** Test structures.

Structure ID	Width nm	Length nm	Series Resistance number of squares
s7dv1/s3dv1	2000	180	5.00
s7dv2/s3dv2	2000	270	5.75
s7dv3/s3dv3	3000	360	5.75
s7dv4/s3dv4	2000	90	4.00
s7dv5/s3dv5	2000	180	4.35
s7dv6/s3dv6	1500	90	4.20
s6dv1	1500	90	4.20
s6dv2	2000	90	3.25
s6dv3	3000	90	2.30
s6dv4	1500	90	5.00
s6dv5	2000	90	4.00
s6dv6	2000	90	4.00

of the electrical channel length, and the formation of unwanted resistive current paths that are beyond the control of the front-gate. Of special interest are the curves with  $V_G=0$  V. Notice that for  $V_D>1$  V,  $I_D$  for  $V_G=0$  V increases with  $V_D$ . With  $V_B=-5$  V,  $I_D$  for  $V_G=0$  V increases sharply for  $V_D>2$  V. This difference in behavior is a result of the applied back-gate bias, which reduces the effective electrical thickness of the channel resulting in stronger front-gate control.

Figure 7.2(b) presents  $I_D-V_G$  curves for a device with a single edge-defined gate, i.e.,  $L_g=90$  nm. The figure illustrates the change in the turn-off behavior of the transistor from the effects of back-gate bias. Note that application of back-gate bias reduces the leakage current and improves the subthreshold slope, resulting in lowered off-state current.

Figure 7.3 presents DC device curves for a representative quadruple-front-gate EdgeFET device, so that  $L_g=4\times90$  nm=360 nm and width  $W_g=3000$  nm. Figure 7.3(b) shows  $I_D-V_G$  curves for this device. The turn-off behavior of this device

(a)  $I_D$ - $V_D$  curves.(b)  $I_D$ - $V_G$  curves at  $V_{DS}=0.3 \text{ V}$ .

**Figure 7.2:** Device curves for a device with  $L_g=90 \text{ nm}$  and  $W_g=2000 \text{ nm}$  at  $\approx 20^\circ \text{C}$ .  $V_s=0 \text{ V}$ .

is similar to that of a device with a single edge-defined gate. The  $I_D$ - $V_D$  curves, however, are more completely saturated, as can be seen in Fig. 7.3(a). This is due to the longer  $L_g$ , which results in decreased DIBL and  $I_D$  for comparable  $V_D$  and  $V_G$ .

Figure 7.4 shows similar device curves reported by Watts and Lee (1993), for a poly-Si TFT with  $L_g=150$  nm. Figure 7.4(a) shows the  $I_D$ - $V_D$  curves, while Fig. 7.4(b) shows the  $I_D$ - $V_G$  curves. Comparison of Fig. 7.4 to the device curves of Fig. 7.2 shows that for  $V_B=-6$  V, devices produced in our work exhibit higher drive current and similar values of subthreshold slope for  $V_B < 0$  V.

### 7.3 Effect of Back-Gate Bias

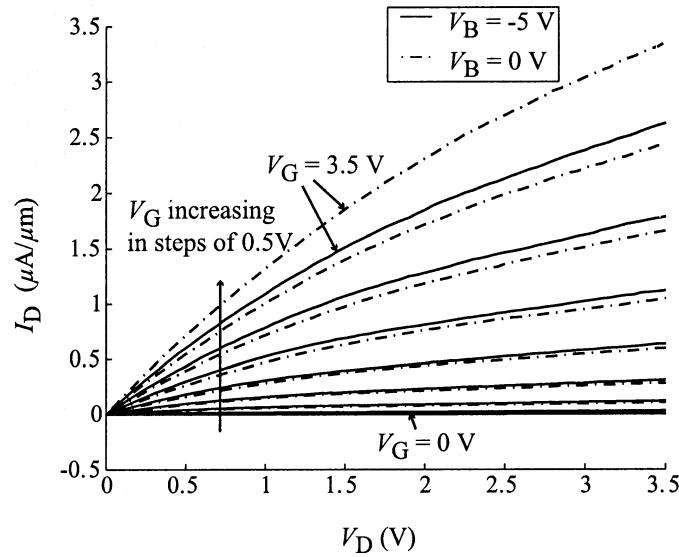
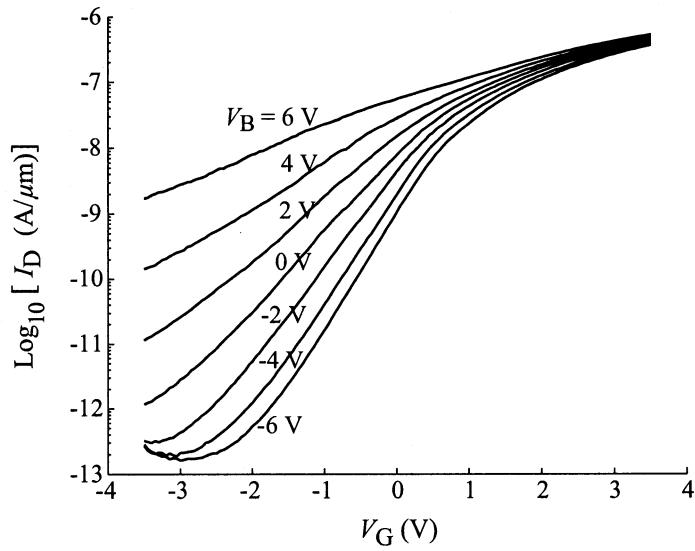
Back-gate bias is a useful tool for adjusting the performance of a transistor using circuit techniques, and hence for optimizing circuit operation for power dissipation at run time. Negative back-gate bias<sup>2</sup> opposes the effect of front-gate potential, and results in electrical channel thickness confinement and reduced device leakage. Positive back-gate bias,<sup>3</sup> on the other hand, assists the front-gate, resulting in faster albeit leakier devices. We report here the effects of the back-gate-bias adjustability on the characteristics of the representative prototype devices.

Figure 7.5 shows the device performance variation of a typical single-gate device ( $L_g=90$  nm,  $W_g=1500$  nm) with the application of back-gate bias,  $V_B$ . Figure 7.5(a) presents the  $I_D$ - $V_G$  family of curves for this device with values of  $V_B$  ranging from  $-6$  V to  $+6$  V. These curves are measured with  $V_D=0.3$  V. Ideally,  $I_D$  should go to zero as  $V_G - V_T < 0$ . This is not the case with real devices, however, where  $I_D$  decreases with decreasing  $V_G$ . In an optimum device,  $dV_G/dI_D \simeq 60$  mV/decade. For bulk CMOS devices,  $I_D$  typically decreases by a decade for about a 100 mV reduction in  $V_G$ . For poly-Si TFTs, the value of subthreshold slope is typically 300–400 mV/decade, which is significantly higher than the ideal value of 60 mV/decade.

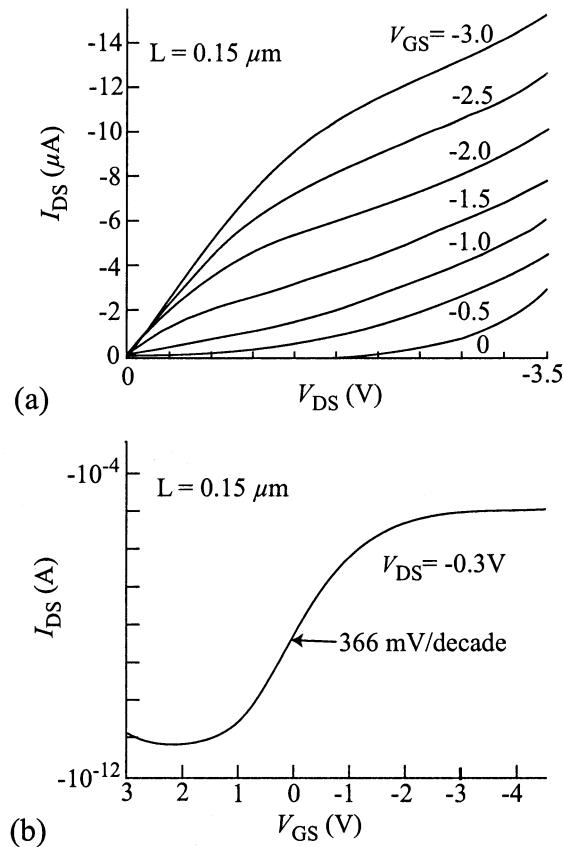
---

<sup>2</sup>Negative back-gate bias with respect to the source, for NMOS transistors, is more commonly known as ‘reverse’ back-gate bias.

<sup>3</sup>Negative back-gate bias with respect to the source, for PMOS transistors, is more commonly known as ‘forward’ back-gate bias.

(a)  $I_D$ - $V_D$  curves.(b)  $I_D$ - $V_G$  curves for  $V_{DS}=0.3 \text{ V}$ .

**Figure 7.3:** Device curves for a device with  $L_g=360 \text{ nm}$  and  $W_g=3000 \text{ nm}$ , at  $\approx 20^\circ \text{C}$ .  $V_s=0 \text{ V}$ .



**Figure 7.4:** Device curves after Watts and Lee (1993). a)  $I_D$ - $V_D$  curves. b)  $I_D$ - $V_G$  curves. Subthreshold slope is about 366 mV/decade. These curves have been reproduced from the digital scan of the original print and enhanced for better resolution and clarity. ©1993 IEEE. Reprinted with permission.

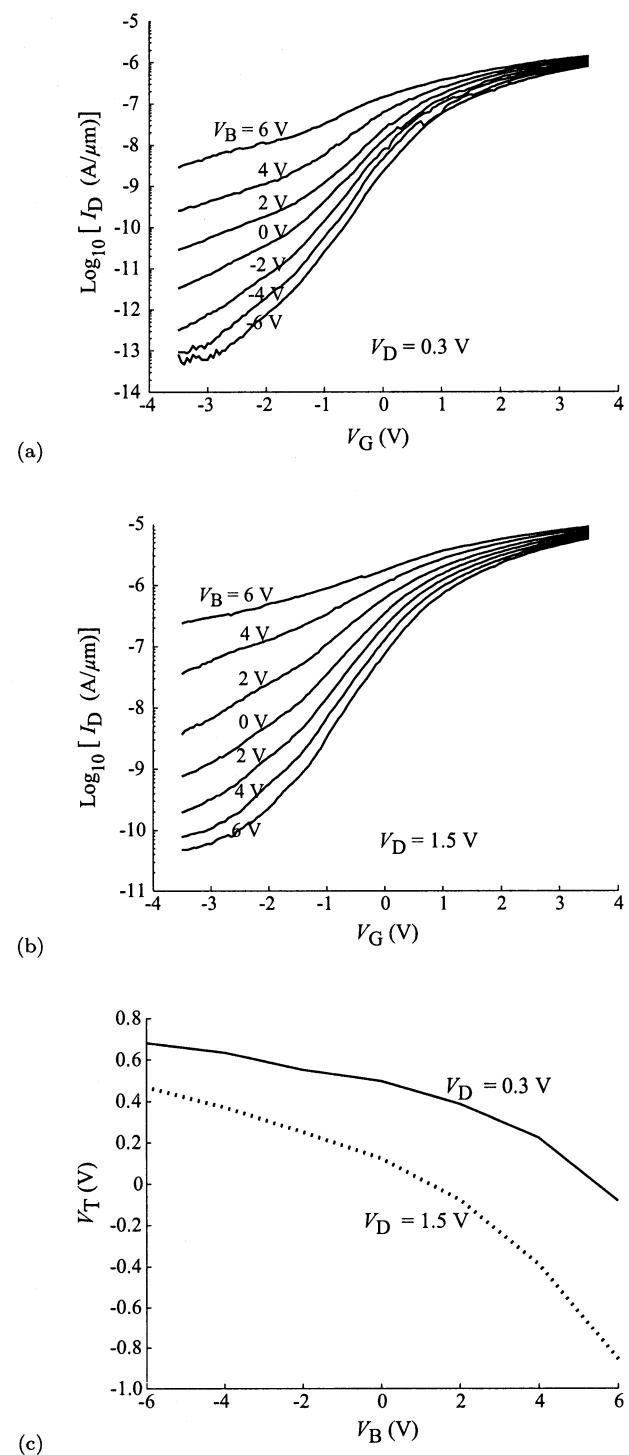
Figure 7.5(a) also shows that the turn-off behavior improves with decreasing values of back-gate bias or increasing reverse bias. Figure 7.5(b) presents the  $I_D-V_G$  family of curves for this device at  $V_D=1.5$  V. Comparison of Fig. 7.5(a) with Fig. 7.5(b) shows that  $I_{\text{off}}$  is 2–3 orders of magnitude higher for  $V_D=1.5$  V. This means that the device turn-off behavior shows dependence on the drain due to DIBL, which results in a lower threshold voltage; hence  $I_D$  increases.

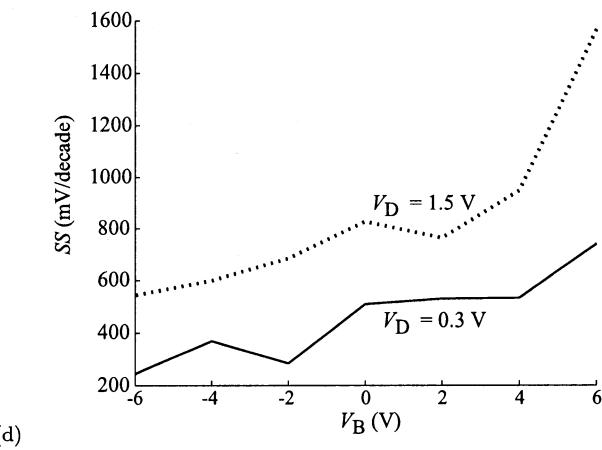
Figures 7.5(c)–(f) show the behavior of four specific parameters vs.  $V_B$ , based on the device  $I_D-V_G$  curves for  $V_D=0.3$  and 1.5 V, described above. Figure 7.5(c) gives the variation of threshold voltage  $V_T$  with back-gate bias  $V_B$ . Figure 7.5(d) shows the subthreshold slope, SS, variation with  $V_B$ . Subthreshold slope decreases with decreasing values of  $V_B$  due to better front-gate control for lower values of  $V_B$ . Figure 7.5(e) shows the variation of the on-state current,  $I_{\text{on}}=I_D|_{(V_G=3.3 \text{ V})}$ , with  $V_B$  for a fixed gate potential of 3.3 V. On-state current decreases with decreasing values of  $V_B$  due to reduced effective channel height from lower values of  $V_B$ . Notice, however, that for  $V_D=1.5$  V,  $I_{\text{on}}$  decreases more rapidly with reduction in  $V_B$  than is the case where  $V_D=0.3$  V. Figure 7.5(f) shows the off-state current,  $I_{\text{off}}=I_D|_{(V_G=0 \text{ V})}$ , variation with  $V_B$  for a fixed front-gate potential of zero volts. Off-state current decreases with decreasing values of  $V_B$ , which reduces the effective channel height. A drain voltage value of 1.5 V results in higher  $I_{\text{off}}$  as compared to the case when  $V_D=0.3$  V, as would be expected.

Figure 7.6 presents the device performance variation of a typical dual-front-gate device (total  $L_g=180$  nm,  $W_g=2000$  nm) with applied back-gate bias. Figure 7.6(a) shows the  $I_D-V_G$  family of curves for this device with values of  $V_B$  from –6 to +6 V; Curves in this group have  $V_D=0.3$  V. Figure 7.6(a) also shows that the turn-off behavior improves with decreasing back-gate bias. Figure 7.6(b) presents the  $I_D-V_G$  family of curves for this device at  $V_D=1.5$  V. Comparison of Fig. 7.6(a) with Fig. 7.6(b) shows that  $I_{\text{off}}$  is 1–2 orders magnitude higher for  $V_D=1.5$  V. This means that the device turn-off behavior shows dependence on the drain voltage, which again may be explained by drain-induced barrier lowering (DIBL).

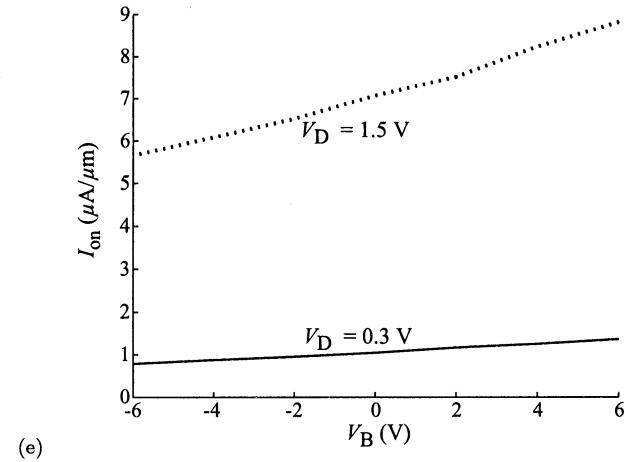
*Next two pages.*

**Figure 7.5:** Effect of applied back-gate bias on the electrical behavior of a transistor with  $L_g=90$  nm,  $W_g=1500$  nm at  $\approx 20$  °C. (a)(b) Effect of back-gate bias on  $I_D-V_G$  curves for  $V_{DS}=0.3$  V, 1.5 V, respectively. (c) Effect of back-gate bias on  $V_T$  at  $V_{DS}=0.3$  V. (d) Effect of back-gate bias on subthreshold slope,  $SS$ , at  $V_{DS}=0.3$  V. (e) Effect of back-gate bias on  $I_{on}$  at  $V_{DS}=0.3$  V. (f) Effect of back-gate bias on  $I_{off}$  at  $V_{DS}=0.3$  V.

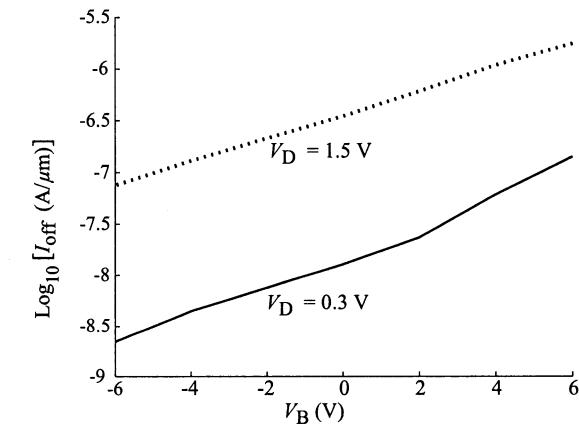




(d)



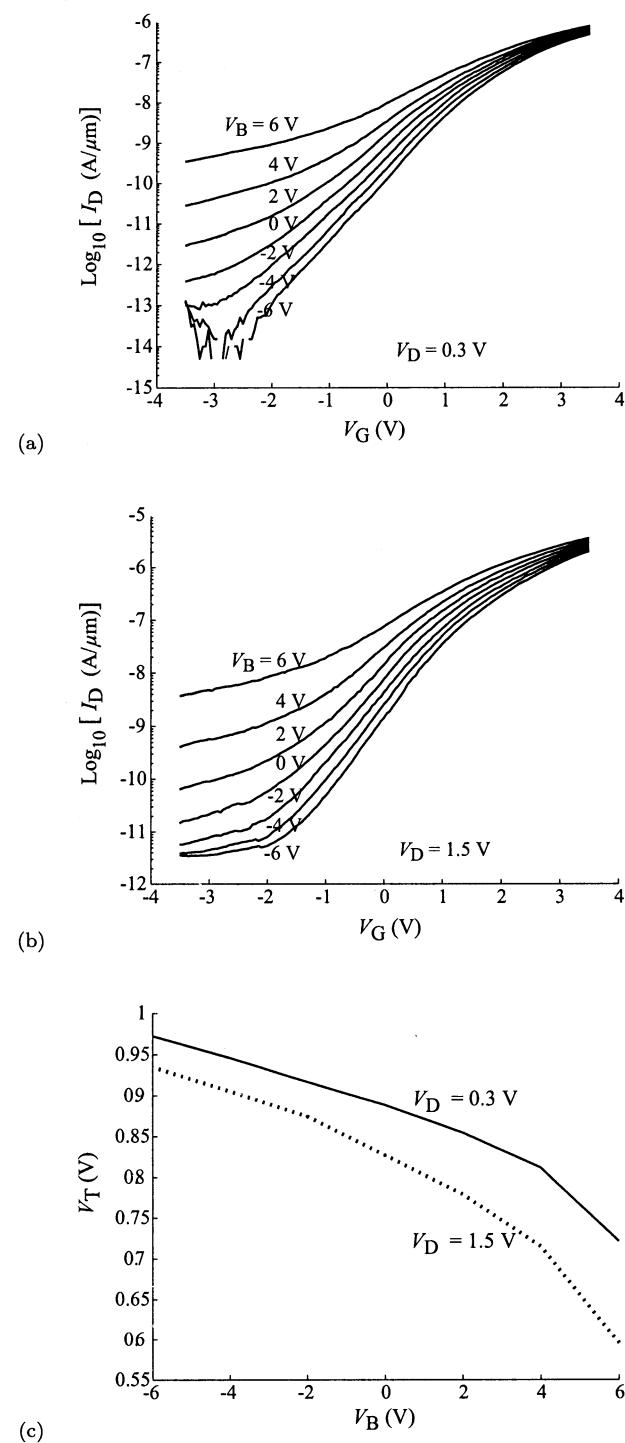
(e)

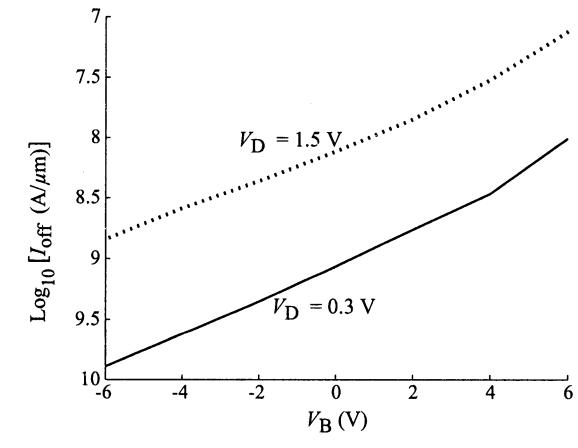
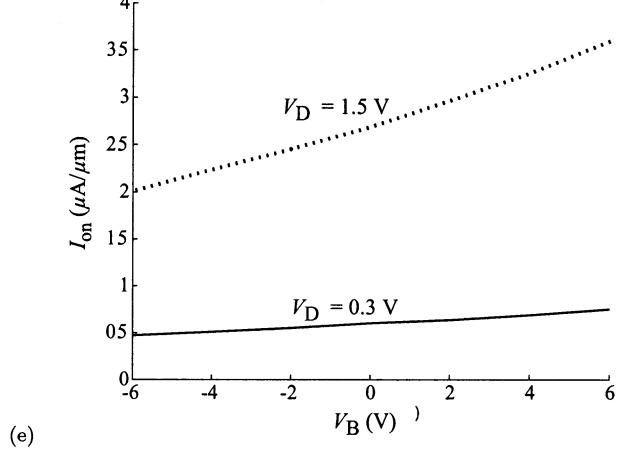
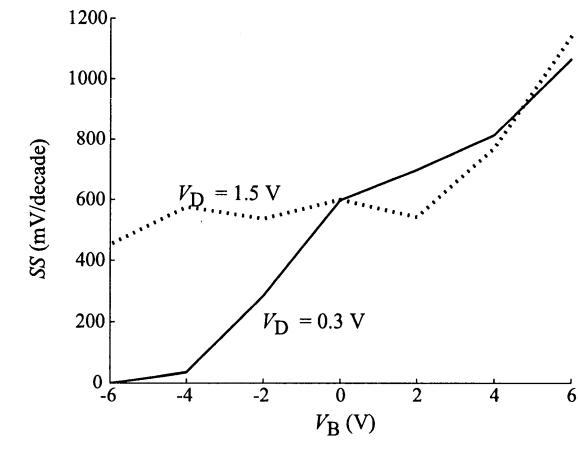


(f)

*Next two pages.*

**Figure 7.6:** Effect of applied back-gate bias on the electrical behavior of a transistor with  $L_g=180$  nm,  $W_g=2000$  nm at  $\approx 20$  °C. (a)(b) Effect of back-gate bias on  $I_D-V_G$  curves for  $V_{DS}=0.3$ , 1.5 V, respectively. (c) Effect of back-gate bias on  $V_T$  at  $V_{DS}=0.3$  V. (d) Effect of back-gate bias on subthreshold slope,  $SS$ , at  $V_{DS}=0.3$  V. (e) Effect of back-gate bias on  $I_{on}$  at  $V_{DS}=0.3$  V. (f) Effect of back-gate bias on  $I_{off}$  at  $V_{DS}=0.3$  V.





Figures 7.6(c)–(f) show the behavior of device parameters vs.  $V_B$ , based on the device  $I_D$ – $V_G$  curves for  $V_D=0.3$  and 1.5 V, described above. Figure 7.6(c) gives the variation of threshold voltage  $V_T$  with back-gate bias  $V_B$ . Threshold voltage increases with decreasing values of  $V_B$ . Figure 7.6(d) shows the subthreshold slope, SS, variation with  $V_B$ . Subthreshold slope decreases with decreasing values of  $V_B$  due to better front-gate control for lower values of  $V_B$ . Figure 7.6(e) shows the variation of  $I_{on}$  with  $V_B$  for a fixed gate voltage of 3.3 V. On-state current decreases with decreasing values of  $V_B$  due to thinner effective channel height from lower values of  $V_B$ . Notice that for  $V_D=1.5$  V, reducing  $V_B$  results in  $I_{on}$  decreasing more rapidly as compared to the case where  $V_D=0.3$  V. Figure 7.6(f) shows the off-state current,  $I_{off}$ , variation with  $V_B$  for a fixed gate potential of zero volts. Off-state current decreases with decreasing values of  $V_B$ , which reduces the effective channel height. A drain-voltage value of 1.5 V results in higher  $I_{off}$  as compared to the case with  $V_D=0.3$  V.

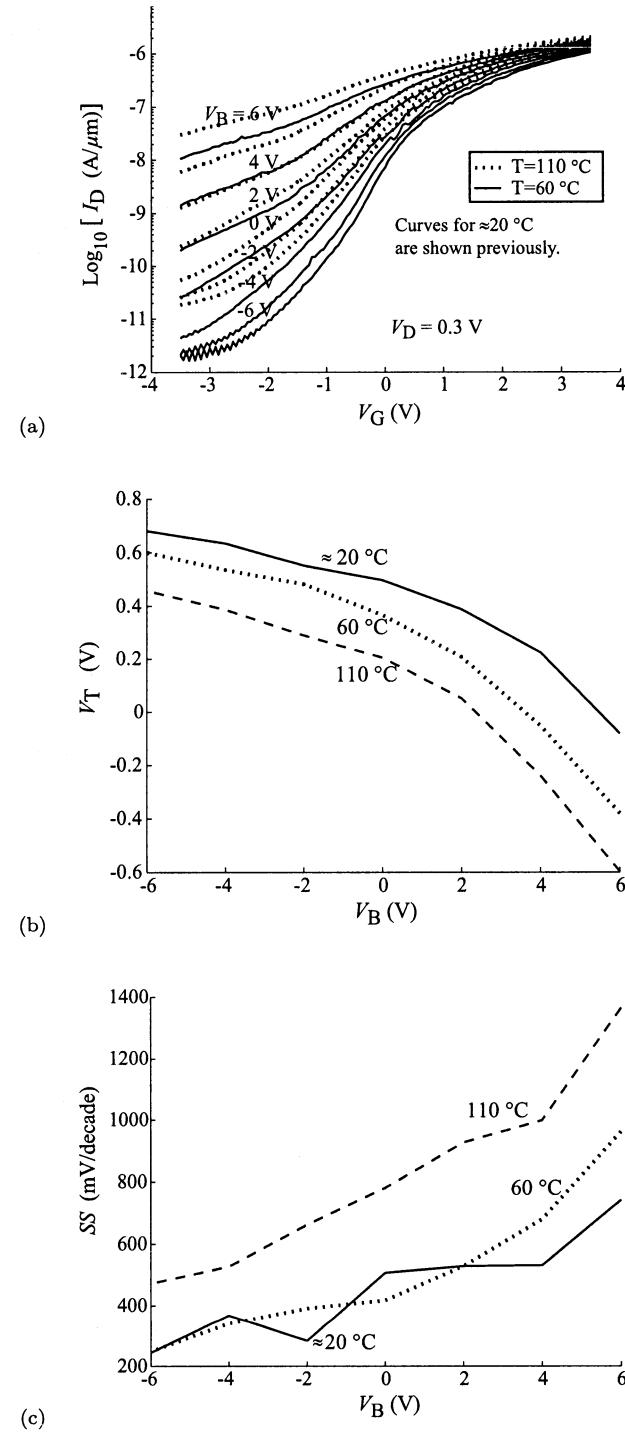
## 7.4 Temperature Dependence of Device Performance

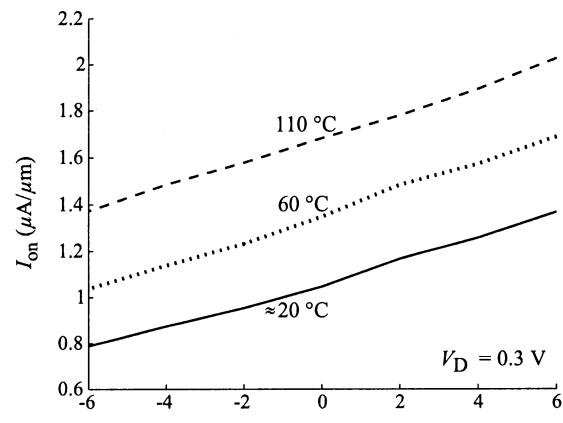
Characterization of the device performance variation over a range of temperatures is important not only to determine the operating range, but also reveals underlying operating phenomena. We therefore extract and present here the device parameters for  $\approx 20$  °C, 60 °C, and 110 °C.

Figure 7.7 presents the device performance variation of a typical single-gate device ( $L_g=90$  nm, and  $W_g=1500$  nm) with applied back-gate bias at  $\approx 20$  °C, 60 °C, and 110 °C. Figure 7.7(a) shows the  $I_D$ – $V_G$  family of curves for this device with values of  $V_B$  from  $-6$  V to  $+6$  V, for  $V_D=0.3$  V at 60 °C and 110 °C. The  $\approx 20$  °C curves have been shown previously in Fig. 7.5. Figure 7.7(a) shows that the turn-off behavior improves with decreasing temperature.

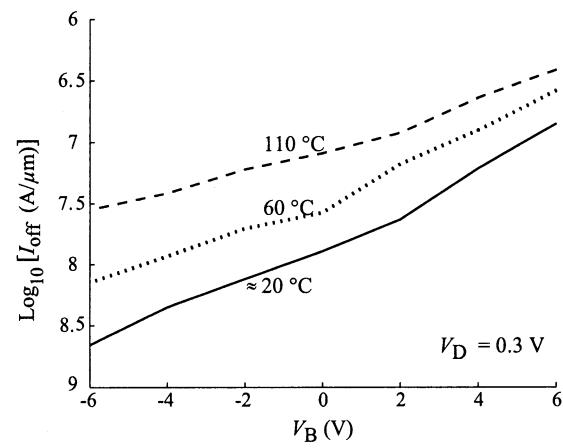
*Next two pages.*

**Figure 7.7:** Effect of temperature on a device with  $L_g=90$  nm,  $W_g=2000$  nm. Curves are plotted for  $\approx 20$  °C, 60 °C, and 110 °C. (a)  $I_D-V_G$  curves at  $V_{DS}=0.3$  V. Please note that this plot only presents curves for 60 °C and 110 °C. Curves for the  $\approx 20$  °C operation of this device are shown in Fig. 7.5. (b)  $V_T$  at  $V_{DS}=0.3$  V. (c)  $SS$  at  $V_{DS}=0.3$  V. (d)  $I_{on}$  at  $V_{DS}=0.3$  V. (e)  $I_{off}$  at  $V_{DS}=0.3$  V. (f)  $I_{on}/I_{off}$  at  $V_{DS}=0.3$  V.

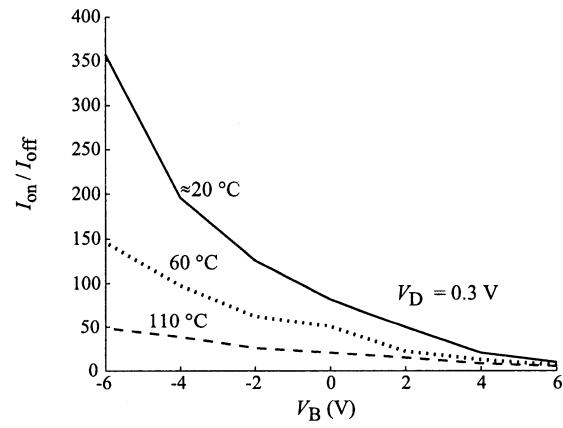




(d)



(e)



(f)

Figures 7.7(b)–(f) show the behavior of several device parameters against  $V_B$ , based on the device  $I_D$ – $V_G$  curves for  $\approx 20^\circ\text{C}$ ,  $60^\circ\text{C}$ , and  $110^\circ\text{C}$ . Figure 7.7(b) shows  $V_T$  variation with  $V_B$ . Higher temperatures result in lower  $V_T$ , as is seen in other CMOS devices.

Figure 7.7(c) shows SS variation with  $V_B$ . SS decreases with lower temperature, which is also consistent with traditional device behavior. Figure 7.7(d) shows  $I_{on}$  variation with  $V_B$ .  $I_{on}$  decreases with lower values of  $V_B$  due to thinner effective channels from lower values of  $V_B$ . Notice, however, that with higher temperature  $I_{on}$  increases, which is unusual. For traditional CMOS devices,  $I_{on}$  decreases with temperature due to a progressive lowering of carrier mobility. In this instance, carrier mobility is already lowered due to poly-Si grain boundaries of the channel; therefore the degradation in mobility with temperature is not apparent in the behavior of  $I_{on}$ . On the other hand, a reduction in  $V_T$  helps to increase  $I_{on}$  with high temperature. A similar but more pronounced trend is obvious for  $I_{off}$ , as seen in Fig. 7.7(e). The ratio  $I_{on}/I_{off}$  is shown in Fig. 7.7(f); there is significant loss in this ratio with the increase in temperature.

## 7.5 Intra-Die Device Performance Variation

Figure 7.8 shows the variation in device performance with back-gate bias for four similar devices located within a few hundred microns radius of one another on a single die. Figure 7.8(a) shows the change in threshold voltage with back-gate bias; note that for device A, e.g.,  $V_T$  can be shifted by approximately 70 mV/V of back-gate bias. Note that devices A, B, and D show similar behavior, however device C has 300 mV and lower threshold voltage compared to the other three devices. This indicates that devices A, B, and C have similar turn-on behavior, whereas device C is clearly different.

Figure 7.8(b) shows the subthreshold slope of all the four devices with applied back-gate bias. All four curves in Fig. 7.8(b) show similar trend, however every

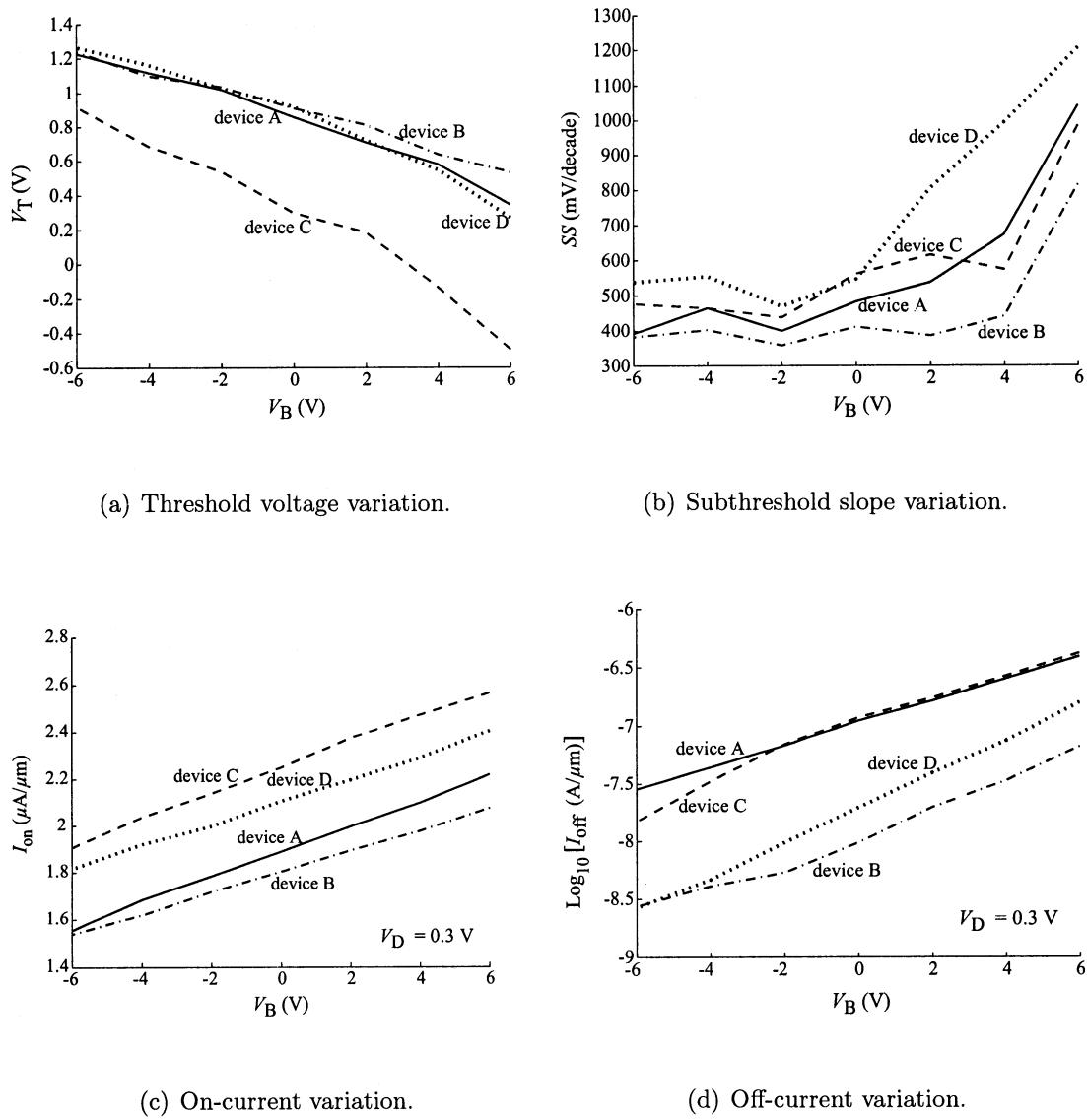
device behaves differently. Device D is the most leaky, whereas device B is the least leaky. This indicates that devices A and C have similar turn-off behavior, whereas devices B and D are different.

Figure 7.8(c) shows the  $I_{on}$  of the four devices against back-gate bias. According to the observations based on Fig. 7.8(a), we expected that devices A, B, and C should show similar  $I_{on}$  curves, and device C should show higher  $I_{on}$ . It is however interesting to see that device D shows higher  $I_{on}$  than devices A and B, even though they show very similar  $V_T$  change with back-gate bias.

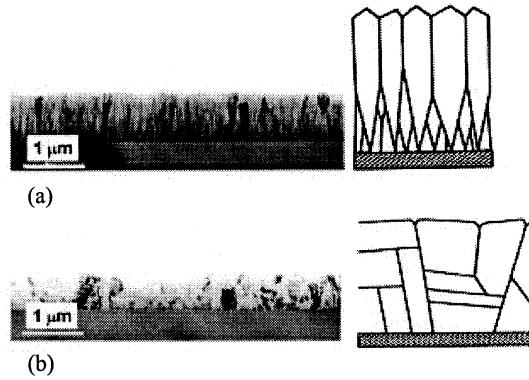
Figure 7.8(d) shows the  $I_{off}$  of the four devices against back-gate bias. Off-state currents of devices A and C are almost identical except for low value of  $V_B$ . Off-state current of device D is higher than device B.

Intra-die variations in the device performance, as observed above, are hypothesized to result from randomly oriented grain boundaries and the enhanced dopant diffusion along grain boundaries. Since the TFT gate length is comparable to grain size, device variations are increased due to less grain-effect averaging as compared with the case of long-channel TFTs where channel has several gains in it. Grain boundaries perpendicular to the channel result in an increase in  $V_T$  and in lowered charge-carrier mobility, thereby increasing subthreshold slope. Those boundaries that are parallel to the channel have the opposite effect, resulting in an increase in subthreshold slope. This behavior is believed to be the result of enhanced source and drain dopant diffusion along the grain boundaries, which results in shortening of the channel length. Similarly, shortening of the channel length in turn increases on-state current as well as off-state current.

Figure 7.9 shows a cross-sectional TEM micrograph comparison of thermally crystallized Si, which was deposited as *a*-Si, with deposited poly-Si (Uma *et al.*, 2001). Figure 7.9(a) shows poly-Si grain structure that appears to be columnar, whereas Fig. 7.9(b) shows grain structure of thermally crystallized Si. It is clear that thermally crystallized Si consists of randomly oriented crystal grains and grain boundaries. Such randomly oriented grain boundaries degrade our device performance.



**Figure 7.8:** Intra-die variations for a device with  $L_g=90 \text{ nm}$ ,  $W_g=2000 \text{ nm}$ . Extracted performance parameters for four devices marked as A, B, C, and D are plotted for comparison. All four devices are located within a few hundred microns radius of one another on a single die.



**Figure 7.9:** Grain structure of poly-Si films observed from cross-sectional TEM. (a) Si film deposited in the polycrystalline form. Grains are organized in columnar fashion. (b) Si film deposited as *a*-Si and later thermally crystallized. Random grain structure is observed in such films. Figure 3 from Uma *et al.* (2001) is reprinted with the author's permission.

It is important to note that in our prototype devices, the gate length of 90 nm is comparable to the crystal grain size achievable in thermally crystallized Si. It therefore appears possible that only a single grain boundary can exist perpendicular to the channel. It is also possible that the grain boundaries in other orientations may also be present. Grain boundaries in different orientations manifest different electrical behavior (Subramanian, 1998). Our observed intra-die device variation results strongly suggest that we are seeing the effects of the presence of very few, possibly only one, grain boundaries in a channel.

In order to address the effects of variations in grain boundary orientation, we examine the effect of grain boundaries along specific planes. Any randomly oriented grain boundary can be projected onto these planes.

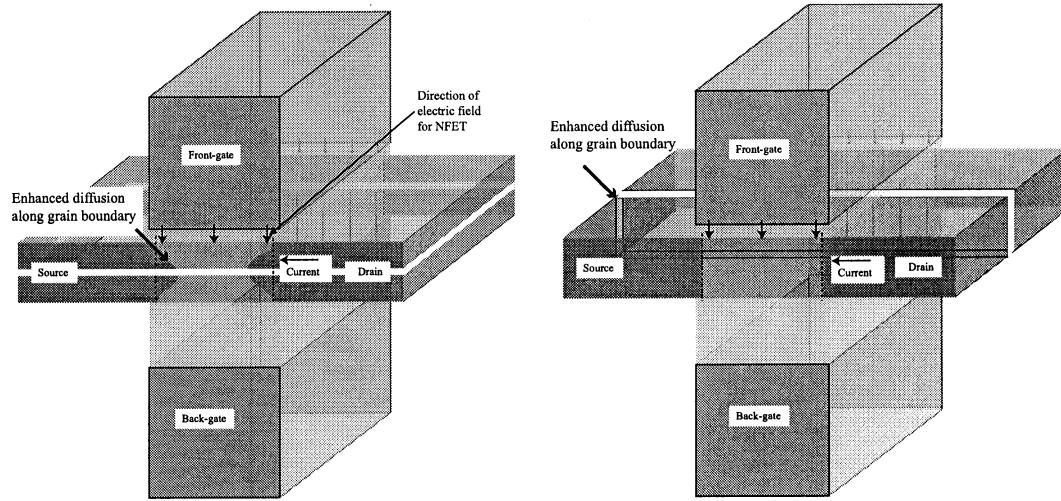
Figure 7.10 shows grain boundaries along different electrical planes with respect to a transistor.

Figure 7.10(a) shows a grain boundary parallel to the plane of the channel resulting in enhanced diffusion of the dopant atoms along the grain boundary, which shortens the channel length. Shortened channel length results in increased leakage, increased on-current, and lowered  $V_T$ . Figure 7.10(b) shows a grain boundary perpendicular to the plane of the channel but parallel to the current flow, resulting in enhanced

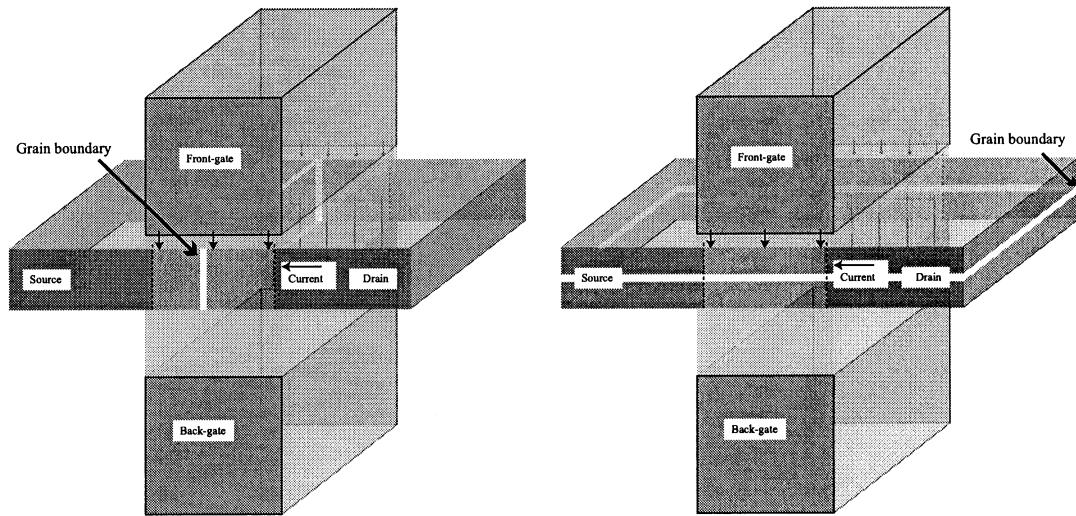
dopant diffusion along the grain boundary that shortens the channel length. It can increases leakage, and can also change on-current.

Figure 7.10(c) shows a grain boundary perpendicular to the current flow in the channel. Such a grain boundary lowers the carrier mobility, and increases  $V_T$  by introducing a higher potential barrier in the channel at the location of the grain boundary. Figure 7.10(d) shows a grain boundary parallel to the plane of the channel. Such a grain boundary lowers carrier mobility, and results in enhanced source and drain dopant diffusion along the grain boundary that shortens the channel length.

Randomly oriented grain boundaries, hence, can change the device behavior in many ways. The effect of the grain boundaries is pronounced in short-channel transistors because the channel length within these transistors is comparable to typical grain size. It is therefore probable that there will be grain boundaries in the channel. By reducing the channel length to be smaller than the average grain size, there is a higher probability of getting fewer grain boundaries in the channel. Our understanding of these phenomenon is based upon the measurements of intra-die device characteristics variation.



- (a) A grain boundary parallel to the plane of the channel results in enhanced diffusion of the dopant atoms along the grain boundary, which shortens the channel length. Shortened channel length results in increased leakage, increased on-current, and lowered  $V_T$ .
- (b) A grain boundary perpendicular to the plane of the channel but parallel to the current flow shortens the channel grain boundary, which increases leakage, and can also affect on-state current.



- (c) A grain boundary perpendicular to the current flow results in lower carrier mobility, and increases  $V_T$  by introducing a higher potential barrier in the channel.
- (d) A grain boundary parallel to the plane of the current flow lowers carrier mobility.

**Figure 7.10:** Effect of grain boundaries on TFT performance.

*We've now gone from generalities to specifics, and from principles to realizable devices which may well someday become practical and mass-produced. So at this point we take stock of where we've been, and wrap up our presentation.*

## Chapter 8

### Summary and Future Work

Technology is never finished. If it were, technologists would need to find other lines of work. But we have grappled in the foregoing chapters with real problems of ULP circuits having ultra-small feature sizes, and shown some paths leading forward into the future. Problems remain, and we also discuss those here. We now present a summary of this work and suggestions for future work.

This dissertation presents the results of a new, “power-aware,” device technology development for ultra-low-power circuits. Practical use of these circuits requires dynamic adjustment of threshold voltages through back-gate bias. A strong concomitant constraint on development of this technology is retaining compatibility with 3-D device integration. Low-power circuits and device technologies are crucial to continuing to extend the trend of Moore’s Law, which postulates that the physical dimensions of semiconductor devices will continue to shrink geometrically with time.

We began this dissertation with a review of the background of low-power design and by describing device requirements for ultra-low-power (ULP) technology. We then presented nanometer-scale considerations for construction and application of thin-film surfaces of Si and SiO<sub>2</sub>. Such films are important building blocks for devices at nanometer-scale range. With this background, we next presented an edge-defined nanolithographic technique, which is a method of nanometer-scale patterning based on use of an ‘edge-definition’ technique that can be implemented with existing lithography equipment. Finally, we described the design, fabrication, and measured

performance results of 90-nm gate-length thin-film transistors fabricated using the edge-defined patterning technique. These transistors feature integral back-gates, and exhibit practical levels of  $V_T$  adjustability with applied back-gate bias. The processing steps required for fabrication of these transistors are compatible with 3-D integration of these devices in stacks of multiple circuit planes.

## 8.1 Summary

Power optimization is important to increase computational-circuit density, especially for 3-D ICs, in order to extend battery life, to reduce environmental impact, and to simplify system design. Microprocessor integrated circuits typically dissipate several tens of watts of power—or more—and contain over 100 million transistors on a small silicon chip. Most integrated circuits execute digital functions, and are commonly implemented with a static-CMOS style of logic. Power dissipation in these circuits results from a combination of transient currents during switching activity and leakage currents. Reducing the supply voltage decreases active switching power as well as leakage power; increasing the threshold voltage also reduces leakage power. Decreasing the supply voltage and increasing the threshold voltage lowers the maximum operating frequency of the circuit, however. A designer may choose static, dynamic, or a combination of static and dynamic power optimization schemes for a particular circuit application. These power-reduction methods require optimization of supply voltages, threshold voltages, operating frequencies, and circuit styles. Supply gating, clock gating, multi- $V_T$  design, multi- $V_{DD}$  design, dynamic voltage scaling, dynamic threshold-voltage scaling, and dynamic frequency scaling are specific techniques used for this purpose. Adiabatic and asynchronous circuit designs are low-power alternatives to static-CMOS circuit style for low-frequency operation.

Our approach to ultra-low-power (ULP) technology is based on the principle of optimizing operation by balancing leakage and switching power for a given frequency of operation. Achieving such a balance requires that the threshold

voltage be adjusted dynamically by varying the back-gate bias. ULP techniques are particularly important—probably critical—for 3-D circuit integration, where heat removal becomes difficult and power-dissipation reduction techniques are required. A device technology for 3-D ULP application follows from fabricating nanometer-scale transistors using nanometer-scale patterning, dynamically adjustable  $V_T$ , and low-temperature processes. Understanding nanometer-scale phenomena and processing techniques is a key to successful development in this effort.

Smooth, thin, Si films are essential for the fabrication of high-performance 3-D devices. Si films thinner than about 20 nm exhibit nucleation-dominated surface roughness. Moreover, the roughness of an underlying  $\text{SiO}_2$  substrate increases the surface roughness of an Si film deposited on top of it. For maximal control of smoothness, thin Si films can be deposited thickly and then ‘etched back.’ Surface cleaning of the substrate prior to Si film deposition does not significantly affect the surface roughness of the deposited *a*-Si film. Ion implantation followed by crystallization smooths the surface, but this is unsuitable for 3-D device stacking due to the requirement for high temperatures. Si films deposited by evaporation from a solid source are smoother than chemically vapor deposited films, although inferior in electrical properties; yet these ‘evaporated’ silicon films remain attractive because of near-room-temperature deposition processing.

Edge-defined lithography can be adapted to permit patterning of nanometer-scale range features using only a micron-scale optical lithography system. In this work, the edge-defined process has been used repeatedly to pattern 180-, 90-, 50-, 25-, and 18-nm structures. Distinctive features of this process are scalability, as demonstrated by 18-nm critical dimension (CD) patterning, a low thermal budget, defect tolerance, and the use of existing processing materials and equipment. CD-SEM measurements provide an evaluation of feature quality. Roughness in the registration sidewall and spacer films contributes directly to CD variations. Further the process is compatible with 3-D stacking of transistor devices. Experimental data suggest that indirect lithography techniques such as the edge-defined process described here potentially

offer cheaper, quicker, and better lithographic solutions at nanometer scales, as compared to conventional lithographic processes.

An ideal device structure for 3-D ULP application comprises an ultra-thin body TFT incorporating a back-gate. We present performance parameters for prototype devices optimized from the system-power-dissipation point of view.

Finally, the processing flow for device fabrication is described. Fabrication of such devices requires deposition of the back-gate, back-gate dielectric, channel, front-gate dielectric, and front-gate. We then describe insulated back-gated TFTs with RTA crystallized poly-Si channels, patterned by use of our edge-defined lithography processing.

The demonstration devices described here show threshold-voltage adjustability with applied back-gate bias across their operating temperature range. Construction of these devices addresses key processing-technology issues for devices suitable for low-power circuit techniques for 3-D device stacking. Threshold-voltage variations are, however, observed in these devices. These variations are plausibly attributed to poly-Si grain-boundary charge traps, and to channel shortening due to increased dopant diffusion along grain boundaries.

## 8.2 Contributions

Contributions of this dissertation to the field of low-power electronics are summarized below:

- We developed an overall transistor fabrication technology framework for electrically-adjustable nanometer-scale transistors that can be used for the implementation of 3-D ultra-low-power circuits.
- We studied deposited films of Si and  $\text{SiO}_2$ , and characterized these for nanometer-scale surface roughness using atomic force microscopy. We developed a method for achieving smooth deposited Si films thinner than 10 nm, using wet etching alternately in nitric acid and hydrofluoric acid.

- We devised a 3-D compatible, low-temperature, edge-defined, nanometer-scale patterning technique, and used it to fabricate structures as small as 18 nm using only standard CMOS photolithography equipment which can directly produce features only larger than 1000 nm.
- We developed and briefly describe novel layout techniques for edge-defined nanolithography capable of reducing mask complexity.
- We designed and optimized demonstration TFT devices for 3-D ULP applications. We presented sensitivity analysis of the design.
- We fabricated and tested 90-nm gate-length TFT devices with a back-gate and a solid-phase crystallized thin-film Si channels. We characterized intra-die variations due to the comparable size of the transistor gate and the crystal grains in the channel; to our knowledge these are the first published results addressing these issues.

## 8.3 Future Work and Improvements

### 8.3.1 Materials

The following future experiments and developments can provide further interesting and useful insight into the material properties that can support further technology enhancements:

- SEM, STM, and TEM-based studies of very thin Si film growth. Such analyses can provide additional information regarding the crystal structure of these films, which can be useful in further transistor optimization.
- Conduct further film surface studies for different processes, such as hot-wire CVD, atomic layer deposition, etc. Such analyses may reveal better alternatives to LPCVD films in terms of film-surface smoothness.
- Repeat the film-surface studies for films deposited on thermally grown  $\text{SiO}_2$  substrates in order to further refine methods and to better define the range of

variations in roughness. It will be interesting to see how very thin films of Si can be deposited on thermally grown  $\text{SiO}_2$ .

- Characterize the electrical properties of films thinner than 20-nm thickness. We did not perform direct electrical characterization of such thin Si films. Therefore, it will be very interesting to characterize their electrical properties.
- Refine smoothing techniques based on electrochemical etching and polishing, and on sputter etching. It is intuitively clear that etching processes based on electrochemical etching and sputter etching should flatten the film surface summits due to enhanced electrical field. Therefore such an processes should be able to smooth out the surface.
- Refine methods for flash annealing for crystallization. Solid-phase crystallization using millisecond ‘flashes’ of light to achieve localized heating can be very attractive.

### 8.3.2 Edge-Defined Lithography

Future research for edge-defined nanolithography should—in our opinion—focus on the following topics:

- Investigation and development of density-enhancement techniques using edge-defined patterning. Such techniques are important in order to integrate more functionality into the same IC area.
- Incorporation of the edge-defined patterning technique to scale features below 10 nm. This is an important topic to explore, since photolithography may not be able to achieve such fine resolutions, and an edge-defined-assisted technique can be of significant value.
- Investigation of binary replication of the edge-defined technique. The edge-defined process can be repeated on the same wafer. The registration polygon patterning can be implemented using an edge-defined lithography, as opposed to photolithography. The edge-defined registration polygon can be used to make a

second layer edge-defined pattern. Structures similar to gratings can be fabricated in this way.

- Investigation of the effects of films smoother than polysilicon or  $\alpha$ -Si for edge-registration, and explore other etching processes. Such an exploration may result in improved CD control.

### 8.3.3 ULP device optimization

The devices demonstrated in this dissertation are not fully optimized, and are mainly limited by the polycrystalline nature of the channel, non-optimal channel thickness, and the gate-oxide thickness and quality. For a 90-nm device, the optimum channel thickness is estimated to be in the range of about 6–9 nm, and the gate-oxide thickness about 2 nm; however, fabricating these films with a smooth Si-SiO<sub>2</sub> interface is difficult, especially in TFTs. Controlled nucleation and lateral crystallization techniques, ozone oxidation for gate oxide, and nickel silicidation on raised sources/drains and gates will improve device performance. With these improvements, these devices should enable high-performance 3-D stacking in VLSI applications.

## 8.4 Concluding Remarks

We have presented edge-defined thin-film transistors that incorporate  $V_T$  adjustability with back-gate bias. These devices are compatible with vertical stacking on Si, or for use over flexible substrates. The adjustability of  $V_T$  with back-gate bias makes these devices a practical technology for ULP circuit applications.

## **Appendices**

# Appendix A

## Relevant Device Physics

### A.1 Mass Action Law

The generation rate of electron-hole pairs in a semiconductor ( $G$ ) is a function ( $f_1$ ) of temperature ( $T$ ), but to first order it is independent of the number of carriers.

$$G = f_1(T) \quad (\text{A.1})$$

**Table A.1:** Constants.

Electron charge	$q$	$1.602 \times 10^{-19} \text{ C}$
Free space permittivity	$\epsilon_0$	$8.854 \times 10^{-14} \text{ F cm}^{-1}$
Speed of light in vacuum	$c$	$2.998 \times 10^{10} \text{ cm s}^{-1}$
Relative permittivity of silicon	$\epsilon_s$	11.7
Relative permittivity of $\text{SiO}_2$	$\epsilon_{\text{ox}}$	3.9
Free electron mass	$m_e$	$9.11 \times 10^{-31} \text{ kg}$
Planck's constant	$h$	$4.135 \times 10^{-15} \text{ eV s}$
Boltzmann's constant	$k$	$8.62 \times 10^{-5} \text{ eV K}^{-1}$
Avogadro's number	$A_o$	$6.022 \times 10^{23} \text{ molecules (g mole)}^{-1}$
Thermal voltage at 300 K	$kT/q$	25.86 mV
Intrinsic carrier density in Si at 300 K	$n_i$	$1.45 \times 10^{10} \text{ cm}^{-3}$

**Table A.2:** Variables.

Free electron concentration	$n$
Free hole concentration	$p$
Acceptor concentration	$N_a$
Donor concentration	$N_d$
Dopant concentration in the body	$N_b$
Depletion charge density	$Q_d$
Electron affinity	$\chi$
Work function	$\phi$
Work function of the gate material	$\phi_M$
Work function of the channel Si	$\phi_S$
Work function difference of the channel and the gate	$\phi_{MS}$
Work function of the material $i$	$\phi_i$
Electric potential	$\psi$
Electric potential in the bulk Si	$\psi_B$
Electric potential on the Si surface	$\psi_S$
Fermi energy level	$E_f$
Lowest energy in conduction band	$E_c$
Highest energy in valence band	$E_v$
Drain voltage	$V_D$
Gate voltage	$V_G$
Source voltage	$V_S$
Back-gate bias	$V_B$
Threshold voltage	$V_T$
Drain-to-source voltage	$V_{DS}$
Gate-to-source voltage	$V_{GS}$
Gate-to-drain voltage	$V_{GD}$
Back-gate-to-source voltage	$V_{BS}$
Flatband voltage	$V_{FB}$
Drain current	$I_D$
Drain-to-source current	$I_{DS}$
Maximum depth of depletion area	$x_{dmax}$
Depth of depletion area near the drain	$x_{dep}$
Front-gate oxide thickness	$t_{ox}$
Effective channel length	$L_{eff}$
Depletion charge density	$Q_d$

**Table A.3:** Silicon dopant impurities.

Acceptors p-type	Donors n-type
Boron	Phosphorous
Indium	Arsenic
$\text{BF}_2$	Antimony

However the recombination rate ( $R$ ) is a function ( $f_2$ ) of the temperature, as well as of the number of free carriers.

$$R = npf_2(T) \quad (\text{A.2})$$

where  $n$  is the number of electrons ( $\text{cm}^{-3}$ ) and  $p$  is the number of holes ( $\text{cm}^{-3}$ ).

At equilibrium

$$np = n_i^2 \quad (\text{A.3})$$

where  $n_i$  is the concentration of free carriers, electrons and holes, in an intrinsic semiconductor material.

$$p = n_i \exp\left(-\frac{q\psi}{kT}\right) \quad (\text{A.4})$$

$$n = n_i \exp\left(-\frac{q\psi}{kT}\right) \quad (\text{A.5})$$

## A.2 Fermi Energy

$\psi(x)$  is defined to be the potential in silicon at an arbitrary depth of  $x$ . The surface of the channel Si is taken as  $x=0$ .

$$\psi(x) = \frac{1}{q} [E_f - E_i(x)] \quad (\text{A.6})$$

$\psi(x)$  is variable along the  $x$ -axis.

$\psi_B$  is the potential inside bulk silicon according to the above mentioned definition. For a p-doped semiconductor,  $\psi_B$  is given by

$$\psi_B = -\frac{kT}{q} \left( \frac{N_a}{n_i} \right) \quad (\text{A.7})$$

where  $N_a$  is the acceptor concentration. For an n-doped semiconductor,  $\psi_B$  is given by

$$\psi_B = -\frac{kT}{q} \left( \frac{N_d}{n_i} \right) \quad (\text{A.8})$$

where  $N_d$  is the donor concentration.

### A.3 Poisson's Equation

Poisson's equation for a silicon region with only the mobile charge (electron) density is given as

$$\frac{d^2\psi}{dx^2} = \frac{q}{\epsilon_s \epsilon_o} n_i e^{q\psi/kT} \quad (\text{A.9})$$

The hole density is negligible for NMOS with  $q\psi/kT \gg 1$ . The silicon channel is assumed to be undoped or lightly doped; therefore  $n_i e^{q\psi/kT} \gg N_b$  i.e. the bulk doping concentration.

### A.4 Threshold Voltage for a Bulk MOS Transistor

A popular definition of the threshold voltage ( $V_T$ ) is

$$V_T = V_{FB} + V_S + 2 |\psi_B| + \frac{|Q_d|}{C_{ox}} \quad (\text{A.10})$$

with depletion charge density ( $Q_d$ ) at threshold

$$Q_d = -qN_a x_{dmax} = -\sqrt{2\epsilon_s q N_a (2|\psi_B| + |V_S - V_B|)} \quad (\text{A.11})$$

where

$$\phi_{MS} = \phi_M - \phi_S \quad (A.12)$$

$$\phi_S = \chi + \frac{E_g}{2q} + \psi_B \quad (A.13)$$

for n+ poly

$$\phi_M = \chi \quad (A.14)$$

for p+ poly

$$\phi_M = \chi + \frac{E_g}{q} \quad (A.15)$$

Therefore, the threshold voltage of a bulk MOS transistor is given by

$$V_{FB} = \phi_{MS} - \frac{Q_f}{C_{ox}} - \underbrace{\frac{1}{C_{ox}} \int_0^{t_{ox}} \frac{x\rho(x)dx}{t_{ox}}}_{\text{Charge in gate oxide}} \quad (A.16)$$

where  $Q_f$  is the fixed charge at the silicon–silicon dioxide interface. The term with underbrace represents charge trapped in the gate oxide.

Assuming that either there is no charge at the silicon–silicon dioxide interface, or else that there is no charge trapped inside the silicon dioxide,

$$V_{FB} = \phi_{MS} \quad (A.17)$$

If source and bulk are grounded,

$$V_T = V_{FB} + 2|\psi_B| + \frac{\sqrt{4\epsilon_s q N_a \psi_B}}{C_{ox}} = V_{FB} + (1 + \frac{6t_{ox}}{x_{dmax}})2\psi_B \quad (A.18)$$

where

$$x_{dmax} = \sqrt{\frac{4\epsilon_s \psi_B}{q N_a}} \quad (A.19)$$

and

$$\frac{\epsilon_s}{\epsilon_{ox}} \approx 3 \quad (A.20)$$

Body effect is

$$\Delta V_T = \gamma \left[ \sqrt{2|\phi_p| + |V_{SB}|} - \sqrt{2|\phi_p|} \right] \quad (\text{A.21})$$

Threshold voltage roll-off for short channel devices is given by Liu et al. (Liu *et al.*, 1993) as

$$\Delta V_T = \frac{2V_x + (V_{DS} + V_x)(1 - e^{-\frac{L_{eff}}{l}}) + 2\sqrt{V_x^2 + V_x(V_x + V_{DS})(e^{\frac{L_{eff}}{l}} - 1)}}{4 \left( \sinh \left( \frac{L_{eff}}{2l} \right) \right)^2} \quad (\text{A.22})$$

where  $V_x = V_{bi} - \psi_S$ ,  $V_{bi}$  is the built-in potential for the source-drain junction,  $\psi_S$  is the surface potential of the inversion layer, and

$$l = \sqrt{\frac{\epsilon_s t_{ox} x_{dep}}{\epsilon_{ox} \eta}} \quad (\text{A.23})$$

where  $x_{dep}$  is the width of the depletion region near the drain.

## A.5 Threshold Voltage of an Asymmetric Double-Gate Thin-Film Transistor

$$V_T \approx \frac{E_g}{2q} + \frac{(\epsilon_s/\epsilon_{ox})t_{ox}}{t_{si} + (\epsilon_s/\epsilon_{ox})2t_{ox}} \Delta\phi_2 + \frac{t_{si} + (\epsilon_s/\epsilon_{ox})t_{ox}}{t_{si} + (\epsilon_s/\epsilon_{ox})2t_{ox}} \Delta\phi_1 \quad (\text{A.24})$$

where

$$\Delta\phi_1 = \phi_1 - \phi_i \quad (\text{A.25})$$

$$\Delta\phi_2 = \phi_2 - \phi_i \quad (\text{A.26})$$

where  $\Delta\phi_i$  is the work function difference between a gate electrode material, #i, and intrinsic silicon.  $\Delta\phi_1$  is for gate material #1, and  $\Delta\phi_2$  is for gate material #2. For a midgap work-function gate,  $\Delta\phi_i = 0$ . For n+ poly gate  $\Delta\phi_i = -E_g/2q$ , and for p+ poly gate,  $\Delta\phi_i = +E_g/2q$  etc. (Taur, 2001).

## A.6 Drain Current of a Long-Channel Bulk MOS Transistor

Drain current in the linear region of a long-channel transistor is given by

$$I_{D1} = \mu_{\text{eff}} \frac{W}{L} C_o x \left[ (V_G - V_T) V_{DS} - \frac{V_{DS}^2}{2} \right] \quad (\text{A.27})$$

Saturation voltage is given by

$$V_{D_{\text{sat}}} = (V_G - V_T) + V_G \quad (\text{A.28})$$

Saturation current is given by

$$I_D = \frac{k}{2} (V_G - V_T)^2 \quad (\text{A.29})$$

Saturation current with channel-length modulation is given by

$$I_D = \frac{k}{2} (V_G - V_T)^2 (1 + \lambda V_{DS}) \quad (\text{A.30})$$

Drain-to-source current in the subthreshold region is given by

$$I_{DS} = \mu_{\text{eff}} C_{\text{ox}} \frac{W}{L} (m - 1) \left( \frac{kT}{q} \right)^2 e^{q(V_G - V_T)/m kT} (1 - e^{-qV_{DS}/kT}) \quad (\text{A.31})$$

## A.7 Drain Current of a Short-Channel Bulk MOS Transistor

$$I_{DS} = v_{\text{sat}} \cdot C_{\text{ox}} \cdot W_{\text{eff}} \cdot (V_{DD} - V_T) \quad (\text{A.32})$$

$$I_{DS} = v_{\text{sat}} \cdot C_{\text{ox}} \cdot W_{\text{eff}} \cdot (V_{DD} - V_T) \cdot \frac{1}{1 + \alpha} \quad (\text{A.33})$$

where  $\alpha$  is

$$\alpha = \frac{E_{\text{sat}} \cdot L_{\text{eff}}}{V_{\text{DD}} - V_{\text{T}}} \quad (\text{A.34})$$

and

$$E_{\text{sat}} = \frac{2v_{\text{sat}}}{\mu_{\text{eff}}} \quad (\text{A.35})$$

$$I_{\text{DS}|V_{\text{DS}} \leq V_{\text{Dsat}}} = \mu_{\text{eff}} C_{\text{ox}} \frac{W_{\text{eff}}}{L_{\text{eff}}} \left( V_{\text{G}'} - \frac{V_{\text{DS}}}{2} \right) V_{\text{DS}} \frac{1}{1 + \frac{V_{\text{DS}}}{E_{\text{sat}} L_{\text{eff}}}} \quad (\text{A.36})$$

$$I_{\text{DS}|V_{\text{DS}} > V_{\text{Dsat}}} = \mu_{\text{eff}} C_{\text{ox}} \frac{W_{\text{eff}}}{L_{\text{eff}} - \Delta l} \left( V_{\text{G}'} - \frac{V_{\text{DS}}}{2} \right) V_{\text{DS}} \frac{1}{1 + \frac{V_{\text{DS}}}{E_{\text{sat}} (L_{\text{eff}} - \Delta l)}} \quad (\text{A.37})$$

where

$$V_{\text{G}'} = V_{\text{GS}} - V_{\text{T}} \quad (\text{A.38})$$

$$V_{\text{Dsat}} = \frac{E_{\text{sat}} L_{\text{eff}} (V_{\text{GS}} - V_{\text{T}})}{E_{\text{sat}} L_{\text{eff}} + (V_{\text{GS}} - V_{\text{T}})} \quad (\text{A.39})$$

$$\Delta l = l \ln \left( \frac{(V_{\text{D}} - V_{\text{Dsat}}) + \sqrt{(V_{\text{D}} - V_{\text{Dsat}})^2 + (E_{\text{sat}} L_{\text{eff}})^2}}{E_{\text{sat}} L_{\text{eff}}} \right) \quad (\text{A.40})$$

## A.8 Off-Current of an MOS Transistor

$$I_{\text{off}} = I_{\text{o}} e^{-\frac{qV_{\text{T}}}{n\text{kT}}} \quad (\text{A.41})$$

where  $I_{\text{o}} \approx n\beta \left( \frac{\text{kT}}{q} \right)^2$ , where  $n$  and  $\beta = \frac{\mu_{\text{eff}} C_{\text{ox}} W_{\text{eff}}}{L_{\text{eff}}}$  are technology constants. A more compact model for  $I_{\text{off}}$

$$I_{\text{off}} = I_{\text{s}} \cdot 10^{-\frac{V_{\text{T}}}{SS}} \quad (\text{A.42})$$

where  $I_{\text{s}}$  is another technology constant and  $SS$  is the subthreshold slope.

## A.9 Power Dissipation in CMOS

Assuming that the short-circuit power is negligible,

$$P_{\text{total}} = P_{\text{active}} + P_{\text{passive}} = aC_{\text{load}} V_{\text{DD}}^2 f + I_{\text{off}} V_{\text{DD}} \quad (\text{A.43})$$

## Appendix B

### Processing Details for Stanford NanoFabrication Facility

In this appendix we include an example of a processing-run sheet for an edge-defined lithography experiment at the Stanford Nanofabrication Facility (SNF). This example covers the basic steps in this process. Please note that as a typical run-sheet, the example sheet makes use of a wafer-numbering system used for one particular experiment. Readers can use any wafer-labeling scheme when duplicating this work.

These procedures are inherently hazardous, and only users with the proper qualifications and training should attempt to carry them out.

The following safety tips always apply:

- Always use a pair of flexigloves over the first pair of gloves before handling cassettes.  
Be particularly careful about glove contamination before and after spin-dry.
- Never mix water into acids; always put acids into water.
- Whenever handling acids, always use a face shield, an apron, and special acid-handling gloves over flexigloves, and gloved over by another flexiglove pair.
- It is easiest to leave the cassettes in place when immersing the cassettes in an acid bath or placing them in the **dump rinse**.
- In a  $\text{H}_2\text{SO}_4:\text{H}_2\text{O}_2$  bath, wafers tend to float out of the cassette. Always watch your cassette for sign of severe bubbling on the photoresist, and agitate/stir as necessary. Sometimes placing another cassette on top helps.

**Table B.1:** Pieces of equipment used at SNF.

Equipment name	Abbreviation	Description
Wet bench nonmetal	WBNonmetal	For wet cleaning. Free of metal contamination.
Wet bench diffusion	WBDiff	For wet cleaning. Free of any contamination.
Wet bench silicide	WBSilicide	For wet cleaning. Metal contaminated.
Wet bench metal	WBMetal	Wet processing for metal films.
Wet bench general	WBGeneral	General, non-specific wet processing.
SVG coater	Svgcoat	Photoresist coating.
Ultratec stepper	Ultratec	Photolithography.
SVG developer	Svgdev	Photoresist developer.
Tylan1-4	Oxide Furnace	Thermal oxidation.
Tylan LPCVD SiO <sub>2</sub>	Tylanbpsg	B, P doped or undoped LPCVD SiO <sub>2</sub> .
Tylan LPCVD Si	Tylanpoly	LPCVD poly-Si, <i>a</i> -Si.
Tylan LPCVD Si	Tystar	LPCVD poly-Si, <i>a</i> -Si with doping.
Innotec evaporator	Innotec	Deposition by ‘evaporation.’
Metal deposition	Gryphon	Deposition by ‘sputtering.’
LAM poly etcher	Lampoly	Poly-Si dry etching.
AME 8100	AMTetcher	SiO <sub>2</sub> , Si <sub>3</sub> N <sub>4</sub> , RIE etching.
AME P500	P5000	Multipurpose RIE etching.
Drytek etcher	Drytek	Multipurpose etching.
AG Associates RTA	RTAAG	Rapid thermal annealing.
Rudolph AutoEL	Rudolph	Ellipsometric thickness measurement.
Nanospec	Nanospec	Reflectance-based film-thickness measurement.
Dimension 3000	Afm2	Atomic force microscope.
Hitachi SEM	SemHitachi	Scanning electron microscope.

- Load and unload cassettes to and from ovens. Warning: Cassettes can become very hot.

**Table B.2:** Processes used at SNF.

Process name	Associated Equipment	Description
Photoresist-stripping	WBNonmetal	Wet photoresist removal.
Pre-LPCVD cleaning	WBDiff/WBmetal	Wet cleaning before LPCVD/oxidation.
Non-diffusion cleaning	WBDiff/WBmetal	Wet cleaning before LPCVD/oxidation.
Dump rinse	All wet benches	Water rinse.
LTO400	Tylanbpsg	LTO deposition at 400 °C.
LTO450P	Tylanbpsg	LTO deposition at 450 °C.
PUMP11	Tylanbpsg	LTO tube idle.
Polyemit	Tylanpoly	Poly-Si deposition at 620 °C.
Poly-Si etching	Lampoly	
Breakthrough etch	Lampoly	
Main etch	Lampoly	
Overetch	Lampoly	
Oxide etching	AMTetcher	

## B.1 Cleaning Procedures

We refer to the photoresist-stripping and pre-LPCVD cleaning procedure several times. Here we outline the details of these cleaning procedures.

The cleaning procedure to remove the photoresist is outlined in Table B.3

**Table B.3:** Photoresist-stripping/non-diffusion cleaning.

Step	Time (min)	H <sub>2</sub> SO <sub>4</sub>	H <sub>2</sub> O <sub>2</sub>	Comments
0	20	90%	10%	120 °C
1	4	-	-	Dump rinse
2	5	-	-	Spin-dry

The standard pre-LPCVD/pre-diffusion cleaning procedure is outlined in Table B.4

**Table B.4: Pre-LPCVD/pre-diffusion cleaning. procedure**

Step	Time (min)	H <sub>2</sub> SO <sub>4</sub>	H <sub>2</sub> O	HCL	HF	H <sub>2</sub> O <sub>2</sub>	Comments
0	10	75%	N/A	-	-	25%	120 °C
1	4	-	-	-	-	-	Dump rinse
2	10	-	70%	15%	-	15%	70 °C
3	4	-	-	-	-	-	Dump rinse
4	0.25	-	98%	-	2%	-	Dip
5	4	-	-	-	-	-	Dump rinse
6	5	-	-	-	-	-	Spin-dry

## B.2 Wafer Scribing

- Use clean tweezers and wafer cassettes.
- Use wafer scriber to scribe wafer numbers on the unpolished side.
- Label each wafer near the wafer bottom flat area. Wafers are labeled 1, 2, 3...25.
  - Wafers 1, 2 = Gate-oxide monitors.
  - Wafers 3, 4 = Gate-poly-Si monitors.
  - Wafers 5, 6 = Stop-oxide monitors.
  - Wafers 7, 8 = Reg poly monitors.
  - Wafers 9–12 = Spacer-oxide monitors.
  - Wafers 13–20 = Device wafers.
- Perform photoresist-stripping on all the wafers at WBNonmetal wetbench in order to remove scribing dust.

### B.3 Gate-Oxide and Gate-Poly-Si Deposition

Gate-poly-Si deposition is carried out right after gate-oxide deposition. The time in between the two procedures should be less than 60 min.

- Perform pre-LPCVD cleaning at WBDiff cleaning station on wafers 1, 2, and 13 through 20.
- The objective is to deposit 3.6-nm oxide on wafers 1, 2, and 13 through 20.
- Load recipe LT0450P in Tylanbpsg.
- LT0450P needs process pressure=250 mT, O<sub>2</sub>=70 sccm, SiH<sub>4</sub>=5 sccm, deposition time  $T_{\text{deposition}} = 00:01:50$ .
- LT0450P has the following important steps:
  - Step 40—Note down the base pressure.
  - Step 45—Note down the leak pressure.
  - Step 65—Hold the process until indicator TEMPC—central zone temperature=450 °C.
  - Step 75—Note down the process pressure.
- Total processing time at Tylanbpsg  $\approx 105 \text{ min} + T_{\text{deposition}}$ , where  $T_{\text{deposition}}$  is based on the recent deposition-rate characterization data for the furnace.
- LT0450P fails leak check if you run this recipe after PSG based recipes. An LT0400 run needs to follow a PSG recipe before LT0450P can be used.
- When LT0450P is at Step 65 and there are about 60 seconds more to go at that step, press the ‘HOLD’ button on the furnace panel to hold the recipe. Now start monitoring TEMPC on the furnace computer. When TEMPC approaches 450 °C, press the ‘HOLD’ button again to resume the recipe. LT0450P needs over 30 min for temperature stabilization, which is longer than the built-in stabilization period; therefore the temperature has to be stabilized manually.
- Very often there is an MFS alarm on the Tylanbpsg furnaces. Whenever this occurs, turn off the alarm sound at the furnace front; however, the alarm stays active. Let the furnace complete its run. After unloading wafers, be prepared to download PUMP11 recipe to the furnace. Press the ‘END’ button at the furnace panel; when

it shows the ‘READY’ message, immediately download PUMP11. This will clear all alarms.

- After unloading wafers, measure thickness on the ellipsometer using the monitor wafer.
- Immediately transfer the remaining wafers to the poly tube for gate poly-Si deposition.
- While loading wafers in Tylanbpsg, start cleaning wafers 3 and 4, the poly-Si thickness monitors. This cleaning procedure should be completed before Tylanbpsg processing completion.
- Poly deposition is performed on wafers 3, 4, and 13 through 20.
- Load recipe **Polyemit** on tylan 9 with target thickness 180 nm, and Deposition time  $T_{\text{deposition}} = 00:19:00$ .
- **Polyemithas** the following important steps,
  - Step 35 - Note down the base Pressure.
  - Step 40- Note down the leak Pressure.
  - Step 65 - Note down the process Pressure.
- Total recipe time = 100 min +  $T_{\text{deposition}}$ .

## B.4 Stop-Oxide and Registration Poly-Si Deposition

- Perform pre-LPCVD cleaning procedure on wafers 5, 6, and 13 through 20 at WBDiff cleaning station.
- Deposit 72-nm LTO on wafers 5, 6, and 13 through 20 using recipe LT0400.
- Load recipe LT0400 in Tylanbpsg.
- LT0400 needs deposition time  $T_{\text{deposition}} \approx 00:03:50$ .
- LT0400P has the following important steps
  - Step 30—Note down the base pressure.

Step 35—Note down the leak pressure.

Step 55—Note down the process pressure.

- Total recipe time  $\approx 80\text{min} + T_{\text{deposition}}$ .
- After unloading wafers, measure the film thickness on the ellipsometer using the monitor wafer.
- Immediately transfer the remaining wafers in the poly tube for poly deposition.
- When you load your wafers in **Tylanbpsg**, start cleaning wafers 7, 8, the registration poly-Si monitors. Their cleaning should be completed before **Tylanbpsg** processing completion.
- Deposit registration poly-Si on wafer 7, 8, and 13 through 20 using recipe **Polyemit**.
- Deposition processing uses recipe **Polyemit** on **Tylanpoly**.
- Deposition target thickness is 216 nm, Deposition time  $T_{\text{deposition}}=00:22:50$
- **Polyemit** has the following important steps:
  - Step 35 - Note Base Pressure.
  - Step 40 - Note Leak Pressure.
  - Step 65 - Note Process Pressure.
- Total recipe time = 100 min +  $T_{\text{deposition}}$

## B.5 Registration Photolithography

- Place wafers 13 through 20 in photo cassettes in lithography area.
- Put the cassette in the oven at @150 °C for 30 min.
- Coat wafers on **Svgcoat**, the automatic photoresist coating system, with 1000-nm thick photoresist, brand name **Shiply3612**.
- Expose the photoresist using the appropriate mask on **Ultratec**.
- Develop the photoresist at **Svgdev** automatic photoresist-developing system.

## B.6 Registration Poly-Si Patterning

- Wafers 13 through 20 go through six dump rinse cycles.
- Perform Lampoly chamber conditioning with a dummy wafer.
- Load poly-Si etching recipe from the disk, and etch the registration layer to pattern it.
- Use 10 sec breakthrough etch; manually stop the main etch when the machine indicates etching completion and 12 sec overetch.
- Photoresist strip wafers 13 through 20 by wet resist strip at WBNonmetal.

**Table B.5:** Lampoly poly-Si etching recipe.

Step	Breakthrough etch	Main etch	Overetch
Pressure	13 mT	10 mT	15 mT
Cl <sub>2</sub>	-	40 sccm	-
HBr	-	100 sccm	50 sccm
O <sub>2</sub>	-	5 sccm	5 sccm
RF Power (Top)	250 W	250 W	250 W
RF Power (Bottom)	40 W	60 W	45 W

## B.7 Spacer Deposition

- Perform pre-LPCVD cleaning at WBDiff cleaning station on wafers 9 through 20.
- The objective is to deposit 180-nm oxide with half-way rotation. The deposition is to be performed in two steps.
- Load recipe LT0450P in Tylanbpsg.
- For the first deposition step, use recipe LT0450P with process pressure=250 mT, O<sub>2</sub>=70 sccm, SiH<sub>4</sub>=5 sccm, Deposition time  $T_{\text{deposition}} = 00:36:00$ .
- Load all the wafers with the wafer-flats facing up. Load device and monitor wafers.
- After completion of the first deposition step, unload one monitor wafer, measure the deposition thickness, and recalculate the time for the second deposition step.
- Rotate all the wafers by 180°.

- Reload the second deposition step recipe.
- After unloading the wafers, measure the thickness on the ellipsometer using the monitor wafer.

## B.8 Oxide-Spacer Etching

- Season AMTetcher by etching dummy wafers.
- Use an oxide monitor wafer to etch oxide on the ATMetcher to calibrate it. Calculate etching time based on the calibrated etching rate.
- Etch the batch of device wafers using `oxide-etching` recipe.

## B.9 Registration Poly-Si Stripping

- Perform poly-Si etching on the device wafers in Lampoly.
- The etching recipe uses 2 sec of breakthrough etching, 1 sec of main etching, and the overetching step is stopped manually when the machine indicates the etching is completed.

## B.10 Oxide-Mask Formation Etching

- Season AMTetcher by etching dummy wafers.
- Use an oxide monitor wafer to etch oxide on the ATMetcher to calibrate it. Calculate etching time based on the calibrated etching rate. The etching time should be targeted to etch 10% additional film thickness in order to account for processing variations.
- Etch the batch of device wafers using `oxide-etching` recipe.

## B.11 Gate-formation etching

- season lampoly.

- Perform poly-Si etching on the device wafers in Lampoly.
- The etching recipe uses 2 sec of breakthrough etching, 1 sec of main etching, and the overetching step is stopped manually when the machine indicates etching completion.

## References

- Burr, James B. 1993 (August). Stanford ultra low power CMOS. *In: Symposium Record, Hot Chips V, Stanford, CA.*
- Burr, James B. 1995a (October). Cryogenic ultra low power CMOS. *Pages 9–11 of: IEEE Symposium on Low Power Electronics, Oct. 1995, San Jose, CA.*
- Burr, James B. 1995b (June). High performance, low-voltage CMOS. *Pages 1.c.1–1.c.75 of: IEEE Computer Elements Workshop, Vail, CO.*
- Burr, James B., & Peterson, A. M. 1991 (October). Energy considerations in multichip-module based multiprocessors. *In: IEEE International Conference on Computer Design: VLSI in Computers and Processors, Cambridge, MA.*
- Burr, James B., & Shott, J. 1994 (February). A 200mV encoder-decoder circuit using Stanford ultra low power CMOS. *In: IEEE International Solid-State Circuits Conference.*
- Burr, James B., Chen, Zongjian, & Baas, B. 1996. *Low-power HF microelectronics: a unified approach.* IEE, UK. Chap. 3.
- Carter, G. 2000. Preferred orientation in thin film growth—the survival of the fastest model. *Vacuum*, **56**, 87–93.
- Choi, Y.-K., King, T.-J., & Hu, C. 2002. A spacer patterning technology for nanoscale CMOS. *IEEE Trans. Electron Devices*, **49**(3), 436–441.

- Claassen, W. A. P., & Bloem, J. 1980. The nucleation of CVD silicon on SiO<sub>2</sub> and Si<sub>3</sub>N<sub>4</sub> substrates. I- The SiH<sub>4</sub>-HCl-H<sub>2</sub> system at high temperatures. *J. Electrochem. Soc.*, **127**(1), 194–202.
- Cobb, Nicolas B. 1998. *Fast optical and process proximity correction algorithms for integrated circuit manufacturing*. Ph.D. Thesis, Univ. of California, Berkeley, California.
- Denton, John P., Pae, Sang Woo, & Neudeck, Gerold W. 2001 (March). Vertical integration of submicron MOSFETs in two separate layers of SOI islands formed by silicon epitaxial lateral overgrowth. *In: GLSVLSI '01*.
- Ditzel, David R. 2000 (August). Transmeta's Crusoe: Low-power x86-compatible microprocessors built with software. *In: Int'l Symp. on Low Power Electronics and Design*.
- Dreier, Troy. 2003. Apple power Mac G5: Neck-and-Neck with Intel PCs. *PC Magazine*, September.
- Flanders, D. C., & Efremow, N. N. 1983. Generation of <50 nm period gratings using edge defined techniques. *J. of Vac. Sci & Tech. (Microelectronics Processing and Phenomena)*, **1**(4), 1105–8.
- Fritzsche, H., & Tsai, C. C. 1979. Porosity and oxidation of amorphous silicon films prepared by evaporation, sputtering and plasma-deposition. *Solar Energy Materials*, **1**(5-6), 471–479.
- Fulks, R. T., Boyce, J. B., Ho, J., Davis, G. A., & Aebi, V. 1999. Laser crystallized polysilicon TFT's using LPCVD, PECVD and PVD silicon channel materials—a comparative study. *Pages 623–628 of: Materials Research Society Symposium—Proceedings*, vol. 557 Warrendale, PA, USA, for Materials Research Society. Also: The 1999 MRS Spring Meeting—Symposium A, ‘Amorphous and Heterogeneous Silicon Thin Films: Fundamentals to Devices,’ 1999, Apr 5–9, San Francisco, CA, USA.

- Horstmann, J. T., Hilleringmann, U., & Goser, K. F. 1998. Matching analysis of deposition defined 50 nm MOSFETs. *IEEE Trans. Electron Devices*, **45**(1), 299–306.
- Hunter, W. R., Holloway, T. C., Chatterjee, P. K., & Tasch, A. F. 1981. A new edge-defined approach for submicrometer MOSFET fabrication. *IEEE Electron Device Letters*, **EDL-2**(1), 4–6.
- Ilhan, Husrev T., Nasrullah, Jawad, Linscott, Ivan, & Tyler, G. Leonard. 2003 (August). Characterization of nano-meter scale roughness of CVD silicon and silicon dioxide films for 3-D device integration. *Pages 539–542 of: Third IEEE Conference on Nanotechnology*.
- Joshi, A. R., & Saraswat, K. C. 2003. High performance submicrometer CMOS with metal induced lateral crystallization of amorphous silicon. *J. Electrochem. Soc.*, **150**(8), 443–449.
- Kamins, T. I. 1980. Structure and properties of LPCVD silicon films. *J. Electrochem. Soc.*, **127**(3), 686–690.
- Kamins, Ted. 1998. *Polycrystalline silicon for integrated circuits and displays*. Norwell, MA: Kluwer Academic Publishers. Chap. 2.
- Kao, D. B., McVittie, J. P., Nix, W. D., & Saraswat, K. C. 1985 (December). Two-dimensional silicon oxidation experiments and theory. *Pages 388–391 of: International Electron Device Meeting*.
- Keshavarzi, Ali, Narendra, Siva, Borkar, Shekhar, Hawkins, Charles, Roy, Kaushik, & De, Vivek. 1999 (August). Technology scaling behavior of optimum reverse body bias for standby leakage reduction in CMOS IC's. *Pages 252–254 of: Int'l Symp. on Low Power Electronics and Design*.
- Lal, Amit, & Blanchard, James. 2004. The daintiest dynamos. *IEEE Spectrum*, September, (NA) 36–41.

- Lee, T. H. 2002. A vertical leap for microchips. *Scientific American*, January.
- Liu, Zhi-Hong, Hu, Chenming, Huang, Jian-Hui, Chan, Tung-Yi, Jeng, Min-Chie, Ko, Ping K., & Cheng, Y. C. 1993. Threshold voltage models for deep-submicrometer MOSFET's. *IEEE transactions on electron devices*, 4(1), 86–95.
- Malhi, S. D. S., *et al.*. 1984. Edge-defined self-alignment of submicrometer overlaid devices. *EDL-5*, 428–9.
- Merkle, Ralph C. 1993. Reversible electronic logic using switches. *Nanotechnology*, 4(1), 21–40.
- Miyasaka, Mitsutoshi, Komatsu, Tadakazu, & Ohshima, Hiroyuki. 1997. Oxidation of amorphous silicon for superior thin film transistors. *Japan J. Appl. Phys.*, 36(4A), 2049–2056. Pt. 1.
- Myers, Chris J. 2001. *Asynchronous circuit design*. Wiley Publishers.
- Nasrullah, Jawad, Burr, James B., & Tyler, G. Leonard. 2003 (August). An edge-defined nano-lithography technique suitable for low thermal budget process and 3-D stackable devices. *Pages 502–505 of: Third IEEE Conference on Nanotechnology*.
- Okazaki, S., Chow, T. P., & Steckl, A. J. 1981. Edge-defined patterning of MOS structures. *IEEE Trans. Electron Devices*, ED-28(11), 1364–1367.
- Saraswat, K. C., Souris, S. J., Banerjee, K., & Kapur, P. 2000. Performance analysis and technology of 3-D ICs. *Pages 85–90 of: Proc. of the 2000 Int'l Workshop on System-Level Interconnect Prediction*.
- Souri, S. J., Banerjee, K., Mehrotra, A., & Saraswat, K. C. 2000. Multiple Si layer ICs: motivation, performance analysis, and design implications. *Pages 213–220 of: Proc. of the 37th Conf. on Design Automation*.

- Strifler, W. A., & Cantos, B. D. 1990. An edge-defined technique for fabricating submicron metal-semiconductor field effect transistor gates. *J. of Vacuum Science & Technology B (Microelectronics Processing and Phenomena)*, **8**(6), 1297–1299.
- Subramanian, Vivek. 1998 (June). *Control of nucleation and grain growth in solid-phase crystallized silicon for high-performance thin film transistors*. Ph.D. Thesis, Stanford University.
- Svilan, V., Matsui, M., Burr, James B., & Sakurai, T. 2000 (August). Energy efficient 32 x 32-bit multiplier in tunable near-zero threshold CMOS. *Pages 268–272 of: Int'l Symp. on Low Power Electronics and Design*.
- Swanson, Richard M. 1975. *Complementary MOS transistors in micropower circuits*. Ph.D. Thesis, Stanford University.
- Takasu, Y., & Todokoro, Y. 1984. Fabrication of submicrometer resist patterns using an edge-defined technique. *Electron. Lett.*, **20**(24), 1013–14.
- Taur, Yuan. 2001. Analytic solutions of charge and capacitance in symmetric and asymmetric double-gate MOSFETs. *IEEE Trans. Electron Devices*, **48**(12), 2861 – 2869.
- To, K. H., & Woo, J. C. 1999. High performance sub- $0.1\mu\text{m}$  SOI polysilicon spacer gate MOSFETs using large angle tilted implant for drain engineering. *Pages 95–99 of: IEEE Int'l SOI Conf. Proc.*
- Uchida, Ken, Koga, Junji, & Takagi, Shin-Ichi. 2003 (December). Experimental study on carrier transport mechanism in double- and single-gate ultrathin-body MOSFETs—Coulomb scattering, volume inversion, and  $\delta T_{SOI}$ -induced scattering. *Pages 805–808 of: International Electron Device Meeting*.

- Uchida, Ken, Koga, Junji, & Takagi, Shin-Ichi. 2004 (June). Carrier transport in ultrathin Si layers of SOI MOSFETs. *Pages 17–18 of: IEEE Silicon Nanoelectronics Workshop, Honolulu, HI.*
- Uma, S., McConnell, A. D., Asheghi, M., Kurabayashi, K., & Goodson, K. E. 2001. Temperature-dependent thermal conductivity of undoped polycrystalline silicon layers. *International Journal of Thermophysics*, **22**(2), 605–616.
- Voutsas, Apostolos T., & Hatalis, Miltiadis K. 1993. Surface treatment effect on the grain size and surface roughness of as-deposited LPCVD polysilicon films. *J. Electrochem. Soc.*, **140**(1), 282–288.
- Watts, R. K., & Lee, J. T. C. 1993. Tenth-micron polysilicon thin-film transistors. *Electron Device Letters*, **14**(11), 515–517.
- Yang, Isabel Y., Vieri, Carlin, Chandrakasan, Anantha, & Antoniadis, Dimitri A. 1997. Back-gated CMOS on SOIAS for dynamic threshold voltage control. *IEEE Transactions on Electron Devices*, **44**(5), 822–831.
- Zhang, S., Han, R., Liu, X., Guan, X., Li, T., & Shang, D. 2001. A novel sub-50nm poly-Si gate patterning technology. *Chinese J. of Semiconductors*, **22**(May), 436–441.