Advanced MOSFET Structures and Processes for Sub-7 nm CMOS Technologies

By

Peng Zheng

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Engineering - Electrical Engineering and Computer Sciences

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Tsu-Jae King Liu, Chair
Professor Laura Waller
Professor Costas J. Spanos
Professor Junqiao Wu

Spring 2016

© Copyright 2016
Peng Zheng
All rights reserved

Abstract

Advanced MOSFET Structures and Processes for Sub-7 nm CMOS Technologies

by

Peng Zheng

Doctor of Philosophy in Engineering - Electrical Engineering and Computer Sciences

University of California, Berkeley

Professor Tsu-Jae King Liu, Chair

The remarkable proliferation of information and communication technology (ICT) – which has had dramatic economic and social impact in our society – has been enabled by the steady advancement of integrated circuit (IC) technology following Moore's Law, which states that the number of components (transistors) on an IC "chip" doubles every two years. Increasing the number of transistors on a chip provides for lower manufacturing cost per component and improved system performance. The virtuous cycle of IC technology advancement (higher transistor density → lower cost / better performance → semiconductor market growth → technology advancement → higher transistor density *etc.*) has been sustained for 50 years. Semiconductor industry experts predict that the pace of increasing transistor density will slow down dramatically in the sub-20 nm (minimum half-pitch) regime. Innovations in transistor design and fabrication processes are needed to address this issue.

The FinFET structure has been widely adopted at the 14/16 nm generation of CMOS technology. Gate-all-around (GAA) FETs are anticipated to be adopted in future generations, to enable ultimate gate-length scaling. This work firstly benchmarks the performance of GAA MOSFETs against that of the FinFETs at 10 nm gate length (anticipated for 4/3 nm CMOS technology). Variability in transistor performance due to systematic and random variations is estimated with the aid of technology computer–aided design (TCAD) three-dimensional (3-D) device simulations, for both device structures. The yield of six-transistor (6-T) SRAM cells implemented with these advanced MOSFET structures is then investigated via a calibrated physically based compact model. The benefits of GAA MOSFET technology for lowering the minimum operating voltage (V_{min}) and area of 6-T SRAM cells to facilitate increased transistor density following Moore's Law are assessed.

In order to achieve similar (or even better) layout area efficiency as a FinFET, a GAA FET must comprise stacked nanowires (NWs), which would add significant fabrication process complexity. This is because stacked NWs are formed by epitaxial growth of relatively thick (>10 nm) $Si_{1-x}Ge_x$ sacrificial layers between Si channel layers to accommodate gate-dielectric/gate-metal/gate-dielectric layers in-between the NWs, so that fin structures with very high aspect ratio (>10:1 height:width) must be etched prior to selective removal of the $Si_{1-x}Ge_x$ layers. Also, it will be more difficult to implement

multiple gate-oxide thicknesses with GAA FET technology for system-on-chip (SoC) applications. In this work, a novel stacked MOSFET design, the inserted-oxide FinFET (iFinFET), is proposed to mitigate these issues. With enhanced performance due to improved electrostatic integrity and minimal added process complexity, iFinFET provides a pathway for future CMOS technology scaling.

Advancements in lithography have been key to sustaining Moore's Law. Due to the low transmittance of blank mask materials and/or the availability of high-intensity light sources for wavelengths shorter than 193 nm, the semiconductor industry has resorted to "multiple-patterning" techniques to increase the density of linear features patterned on a chip. The additional cost due to extra lithography or deposition and etch processes associated with multiple-patterning techniques threaten to bring Moore's Law to an end, stunting the growth of the entire ICT industry. This work proposes an innovative cost-efficient patterning method via tilted ion implantation (TII) for achieving sub-lithographic features and/or doubling the density of features, one that is capable of achieving arbitrarily small feature size, self-aligned to pre-existing features on the surface. The proposed technique can be used to pattern IC layers in both front-end-of-line (FEOL) and low-temperature back-end-of-line (BEOL) processes. With feature size below 10 nm experimentally demonstrated, TII-enhanced patterning offers a cost-effective pathway to extend the era of Moore's Law.

The primary reason for increasing the number of components per IC, enabled by advancement of IC manufacturing technology, was (and still) is lower cost. Although different opinions are held throughout industry regarding the "cost-per-transistor" trend, reduction in IC manufacturing cost is the key challenge as technology advances to extend Moore's Law. This work summarizes a survey regarding IC manufacturing cost throughout the semiconductor industry. Two case studies reveal that the iFinFET technology and TII double patterning technique have significant economic merit in future technology nodes, especially beyond the 7 nm technology node where the industry does not yet have clear solutions. The proposed technologies can enable the semiconductor industry to extend the era of Moore's Law, with broad economic and social benefit to society.

To my family,
for their unbounded love and unwavering support

Table of contents

Abstract 1
Table of contents ii
Acknowledgments iv
Chapter 1: Introduction 1
1.1 Moore's Law: A Historical Perspective 1
1.2 Advanced MOSFET Structures 3
1.3 Advanced Patterning Technologies 4
1.4 Research Objectives and Thesis Overview 5
1.5 References 7
Chapter 2: 10 nm L_g Multi-Gate MOSFET Technologies 9
2.1 Introduction 9
2.2 GAA and SOI FinFET Design Optimization 9
2.3 Impact of Process-Induced Variations 11
2.4 6-T SRAM Cell Performance and Yield 15
2.5 Summary 19
2.6 References 19
Chapter 3: Inserted-Oxide FinFET (iFinFET) 23
3.1 Introduction 23
3.2 iFinFET Structure and Fabrication 23
3.3 Comparison of FinFET, iFinFET and GAA MOSFET Performance Characteristics ... 26
3.4 Impact of Process-Induced Variations 31
3.5 Summary 34
3.6 References 34
Chapter 4: Sub-lithographic Patterning via Tilted Ion Implantation (TII) 36
4.1 Introduction 36

4.2 TII-Enhanced Patterning Concept 37
4.3 Demonstration of Pitch-halving via TII Patterning 40
4.4 TII Patterning Resolution Limit 40
4.5 Summary 50
4.6 References 50
Chapter 5: IC Manufacturing Cost Analyses 52
5.1 Introduction 52
5.2 Manufacturing Cost Challenges for Advanced CMOS Technologies 53
5.3 Manufacturing Cost Comparison of FinFET *vs.* iFinFET *vs.* GAA MOSFET 55
5.4 Manufacturing Cost Comparison of SADP *vs.* TII Double Patterning 57
5.5 Summary 58
5.6 References 59
Chapter 6: Conclusion 60
6.1 Contributions of This Work 60
6.2 Suggestions for Future Work 61
6.3 References 61

Acknowledgments

I would like to express my sincere gratitude to all the people who have helped me during my 3 years and 8 months Ph.D. study at UC Berkeley.

First and foremost, I would like to thank my advisor, Professor Tsu-Jae King Liu, for all the guidance and support she has provided throughout my endeavor. Professor King epitomizes an ideal Ph.D. mentor. It is truly a privilege and great honor for me to learn semiconductor devices and technologies from such a world-renowned expert. Her insights and visions lead me to the forefront of my research field. Her enthusiasm and encouragement keep me confident and motivated when discovering truth. My Ph.D. study at UC Berkeley coincides with Professor King's tenure as department chair. I want to particularly thank her for always making herself available whenever I need help and advice. Her professionalism and character will no doubt have a significantly beneficial impact on my future career. It's truly a rewarding experience for me to work as her student, research assistant and teaching assistant. Thanks very much for everything, Professor King!

I would also like to thank my dissertation committee and my qualifying exam committee. Professor Laura Waller generously chairs my qualifying exam. Her kind suggestions and help have always been an encouragement. I want to thank Professor Costas Spanos for bringing his invaluable insights into the semiconductor industry. It's his great suggestion that the research regarding IC manufacturing cost is conducted in this work. I absolutely benefit a lot from this experience by looking at this industry from a different but very important perspective. I also want to thank Professor Junqiao Wu for his kind suggestions and continuous support.

This work would not have been possible without my research collaborators. Dr. Yi-Bo Liao (now at TSMC) worked closely with me in my first project, although we were physically 15 time-zones apart from each other. Our late-night Skype meetings were always fruitful, and definitely played an important role in helping us complete the project in a timely manner. I thank him for his patience and commitment. Coming from Synopsys, Dr. Daniel Connelly elevated my research skills to a new level. Dan not only taught me how a world top level TCAD engineer work, but more importantly helped me become a better researcher in each single aspect. I thank him for his knowledge, suggestions, questions and discussions. I also want to thank Dr. Sang Wan Kim, Dr. Kimihiko Kato, and Suchit Bhattarai for their help in the TII pattering project. Their hard work and collaboration have always inspired me to push my limit to discover more in this project. Great thanks also go to Dr. Leonard Rubin (Axcelis) and Dr. Bich-Yen Nguyen (Soitec) for bringing resources from the industry, and also Dr. John Dukovic (AMAT), Dr. Ludovic Godet (AMAT), and Dr. Jun Xue (AMAT) for early collaboration of this project. I also want to thank Fei Ding for his assistance (especially when I was facing multiple deadlines). His questions always make me think how I can explain things clearly to a young researcher and how to become a better senior student in the group. Dr. Nattapol Damrongplasit (now at Intel) and Professor Meng-Hsueh (NCKU) have also provided great help in my research.

Most of the experimental work in this dissertation was performed at the UC Berkeley Marvell Nanofabrication laboratory. I would like to acknowledge all the

Nanolab staff for keeping the tools up and providing guidance to overcome process problems. In particular, I thank Dr. William Flounders, Dr. Jeffrey Clarkson, Kim Chan, and Richelieu Hemphill. I am also grateful to Dr. I-Ru "Tim" Chen for the early training in the Nanolab, and also to Dr. Yuping Zeng, Xu "Chris" Zhao (Crossbar), and Zeying Ren (Crossbar) for sharing their knowledge in tool operations. The simulation work in this dissertation would not have been done without accessing the latest version of software and computationally powerful servers (especially my baby venus@eecs). Credit is due to Lars Rohrbach, Kevin Mullally and the whole CUSG team. I would also like to thank Charlotte Jones, Shirley Salanio, Meg Pressley, Sim Kallan, and Lydia Yoon for administrative support.

I am fortunate to have been working with so many great minds of King Group, and those of the wider UC Berkeley Device Group. Many of them helped me professionally and/or personally. In particular, I would like to thank Dr. Byron Ho for being my big brother in my first semester at UC Berkeley, Dr. Nuo Xu for discussions and challenges during my early research stage. I also want to thank Professor Woo Yong Choi, Professor Yi Zhao, Dr. Xin Sun, Dr. Rhesa Nathanael, Dr. Sung Hwan Kim, Dr. Eung Seok Park, Dr. Wook Hyun Kwon, Dr. Jack Yaung, Dr. Yenhao "Philip" Chen, Dr. Alexis Peschot, Dr. Chun Wing Yeung, Dr. Angada Sachid, Dr. Sourabh Khandelwal, Jun Fujiki, Yi-Ting "Dennis" Wu, Chuang Qian, Xi Zhang, Benjamin Osoba, Urmita Sikder, Alice Ye, and Leyla Shams. Special thanks must go to my peers in the device group: Dominic Labanowski, Varun Mishra, Juan Pablo Duarte, Kevin Chen, Thomas Rembert, and Maria Pace. It's pretty much the Sunday practices with you in the last month before the prelim that made me capable of passing the exam.

Life at Berkeley as a student has been such an enjoyable experience thanks to the following individuals. I thank Professor David Allstot, Professor Jeffrey Bokor, Professor Ali Javey, Professor Bora Nikolic, Professor Clark Nguyen, Professor Vladimir Stojanovic and Professor Vivek Subramanian for their wonderful lectures. I also thank the "residents" in 373 Cory Hall for make this room the most memorable to me across campus: Dr. Wei-Chang Li, Dr. Robert Schneider, Dr. Henry Barrow, Dr. Lingqi Wu, Dr. Thura Naing, Ruonan Liu, Alper Ozgurluk, Jalal Naghsh Nilchi, Yafei Li, Yongjun Li, Ricky Liou, Jikang Chen and Esther Chang. I must also thank Yadi Wang, Dr. Zhaoyi Kang, Dr. Bo Zheng and Travis Massey for the help they provided when I first arrived in California.

My two summer internships definitely have broadened my perspectives. The knowledge I've leant from my industry fellows are invaluable to my future career. Sincere appreciation goes to all the people I worked with at Intermolecular and Oracle.

I must also gratefully acknowledge the education I received at Georgia Tech, where my overseas student life first started. I sincere appreciate the guidance and advice from each of these individuals: Professor James D. Meindl, Professor Azad Naeemi, Professor Ajeet Rohatgi, Professor Muhannad Bakir, Professor Oliver Brand, Professor Tomas Gaylord, Professor Sung Ha Kang, Professor Sung Kyu Lim, Professor Saibal Mukhopadhyay, Professor Shyh-Chiang Shen, Dr. Raghu Murali, Dr. Yinxiao "Ian" Yang, Dr. Sarah Bryan, Dr. Kevin Brenner, TJ Beck and all the staff member at MiRC and NRC. Special thanks go to Professor Guotong Zhou for all of her great help (especially for bringing GT into my life).

I am extremely blessed and thankful to my family for their unbounded love and unconditional support. I am deeply indebted to my parents, for all the sacrifices they have made for me and the faith they have in me. They are the rock in my life, and I thank them for who they are. I am also deeply thankful to Minmin's presence in my life. As my best friend, she went through all the ups and downs with me during this Ph.D. journey. I thank her for her love, understanding and encouragement. I look forward to exploring the future adventures of my life, with her as my copilot.

Finally, the support of this work by the EECS department fellowship, Lam Research Fellowship, Impact+ Program, Applied Materials, and NSF Award 0939514 is gratefully acknowledged.

Chapter 1

Introduction

1.1 Moore's Law: A Historical Perspective

The remarkable proliferation of information and communication technology (ICT) – which has had dramatic economic and social impact in our society – has been enabled by the steady advancement of integrated circuit (IC) technology following Moore's Law [1.1-1.2]. In 1965, Dr. Gordon Moore observed and projected that the number of components (transistors) on an IC "chip" doubles every year (**Fig. 1.1** (**a**)). In 1975, as the rate of growth began to slow down, Dr. Moore revised the forecast to doubling approximately every two years (**Fig. 1.1** (**b**)).

Fig. 1.1 (a) Moore's 1965 paper forecasted yearly doubling of the number of components per function [1.1]. (b) Moore's 1975 revision to doubling every two years [1.2].

Increasing the number of transistors on a chip provides for lower manufacturing cost per component (**Fig. 1.2**) and improved system performance. The virtuous cycle of IC technology advancement (higher transistor density → lower cost / better performance → semiconductor market growth → technology advancement → higher transistor density *etc.*) has been sustained for 50 years (**Fig. 1.3**).

Fig. 1.2 (a) Moore's 1965 statement regarding minimum manufacturing cost reduction with technology advancement over time [1.1]. (b) Intel's cost-per-transistor observation and projection [1.3], Source: Intel Corp.

Fig. 1.3 The number of transistors per chip area has continued to increase (adapted from [1.4]).

1.2 Advanced MOSFET Structures

As Tri-gate MOSFETs (also known as FinFETs) already have been in high volume production since the 22 nm CMOS technology node [1.5-1.8], it is generally believed that multi-gate structures will be necessary to scale the MOSFET gate length down to 10 nm and below.

2003	2005	2007	2009	2011
90 nm	65 nm	45 nm	32 nm	22 nm
SiGe Strained Silicon	2nd Gen. SiGe Strained Silicon	Gate-Last High-k Metal Gate	2nd Gen. Gate-Last High-k Metal Gate	Tri-Gate

Fig. 1.4 Tri-gate MOSFETs (also known as FinFETs) have been in high volume production since the 22 nm CMOS technology node (adapted from [1.9]).

FinFETs require forming high aspect ratio Si stripes (narrow stripe width to maintain good electrostatic integrity and tall stripe height to achieve high current per unit layout area). To enable ultimate transistor gate length scaling, stacked gate-all-around (GAA) MOSFETs is anticipated to be adopted eventually in future CMOS technology generations [1.10-1.11].

(a) 3D-NWFET (b) ΦFET

Fig. 1.5 Cross-sectional TEM images of stacked GAA MOSFETs [1.11].

1.3 Advanced Patterning Technologies

Advancements in lithography have been key to sustaining Moore's Law. As CMOS technologies advance, however, lithography continues to be challenged to define ever-shrinking feature sizes and density. Due to the low transmittance of blank mask materials and/or the availability of high-intensity light sources for wavelengths shorter than 193 nm [1.12], the semiconductor industry has resorted to "multiple-patterning" techniques to increase the density of linear features patterned on a chip [1.10]. Spacer lithography [1.13], also known as self-aligned double patterning (SADP) [1.14], has been the workhorse in high-volume manufacturing since the 22 nm technology node. **Fig. 1.6** illustrates this double-patterning technique.

(a) SACRIFICIAL LAYER
IC LAYER
(b) PHOTORESIST
(c)
(d)
(e)
(f)
(g)
(h)
(i)
(j)

Fig. 1.6 Schematic cross-sections illustrating the self-aligned double patterning technique: (a) the IC layer to be patterned is coated with a sacrificial layer; (b) photoresist is coated onto the sacrificial layer; (c) photolithography is used to print features in the photoresist layer; (d) an etch process is used to remove regions of the sacrificial layer in regions not protected by the photoresist; (e) photoresist is selectively removed; (f) a relatively thin hard mask layer is conformally deposited; (g) an anisotropic etch process is used to form hard-mask "spacers" along the sidewalls of the sacrificial layer patterns (note that the width of these spacers is correlated with the thickness of the deposited hard mask layer, and can be much smaller than the lithographic resolution limit); (h) the sacrificial layer is selectively removed; (i) an etch process is used to remove regions of the IC layer not protected by the spacers; (j) the spacers are selectively removed.

Another commonly used double-patterning technique is referred to as "double exposure, double etch" or "litho-etch-litho-etch (LELE)" and is illustrated in **Fig. 1.7**. As implied by its name, it involves roughly twice the number of processes as the conventional process to pattern a single IC layer.

Fig. 1.7 Schematic cross-sections illustrating the "double exposure double etch" technique: (a) the IC layer to be patterned is coated with a hard mask layer; (b) a first layer of light-sensitive "photoresist" is coated onto the hard mask layer; (c) photolithography (light exposure through a mask, followed by immersion in a chemical developer solution to remove photoresist in regions exposed to light) is used to print features in the photoresist layer (note that these features usually are "trimmed" to become narrower than the lithographic resolution limit); (d) an etch process is used to remove regions of the hard mask layer in regions not protected by the etch-resistant photoresist; (e) photoresist is selectively removed; (f) a second layer of photoresist is coated;(g) photolithography is used to print features in the photoresist layer – inevitably misaligned with the features defined by the 1st photoresist layer; (h) an etch process is used to remove regions of the IC layer not protected by either the hard mask or photoresist; (i) hard mask and photoresist layers are selectively removed.

Since extreme ultraviolet (EUV) lithography won't be ready for the early stages of high-volume manufacturing (HVM) at the 7 nm technology node [1.15], the industry expects to use "self-aligned quadruple patterning (SAQP)" and/or "litho-etch-litho-etch-litho-etch (LELELE)" multiple-patterning techniques based on 193 nm immersion lithography to achieve the desired smaller feature sizes [1.10]. However, the additional cost due to extra lithography or deposition and etch processes (each of which involve multiple steps, e.g. anti-reflection coating, bake, pre-clean, etc.) associated with multiple-patterning techniques threatens to increase cost-per-transistor and hence may bring Moore's Law to an end, stunting the growth of the entire ICT industry.

1.4 Research Objectives and Thesis Overview

Semiconductor industry experts predict that the pace of increasing transistor density will slow down dramatically in the sub-7 nm (minimum half-pitch) regime. Innovations in transistor design and fabrication processes are thus urgently needed to address this issue. This work aims to enable the semiconductor industry to extend the era of Moore's Law, with broad economic and social benefit to society.

In **Chapter 2**, the performance of GAA MOSFETs is benchmarked against that of optimized SOI FinFETs at 10 nm gate length (anticipated for 4/3 nm CMOS technology). Variability in transistor performance due to systematic and random variations is estimated with the aid of TCAD 3-D device simulations, for both device structures. The yield of 6-T SRAM cells implemented with these advanced MOSFET structures is then investigated via a calibrated physically based compact model. GAA MOSFET technology is projected to provide for 0.1 V lower minimum cell operating voltage with reduced cell area.

In order to achieve similar (or even better) layout area efficiency as a FinFET, a GAA FET must comprise stacked nanowires (NWs), which would add significant fabrication process complexity. Also, it will be more difficult to implement multiple gate-oxide thicknesses with GAA FET technology for system-on-chip (SoC) applications. In **Chapter 3**, a novel stacked MOSFET design, the inserted-oxide FinFET (iFinFET), is proposed to mitigate these issues. The performance of iFinFET is benchmarked against that of the conventional bulk FinFET and stacked-nanowire gate-all-around (GAA) FET, via 3-D device simulations, for both n-channel and p-channel transistors. The results show that the iFinFET provides for improved electrostatic integrity relative to the FinFET, but with substantially less gate capacitance penalty relative to the GAA FET. Thus, iFinFET technology offers a technological pathway for continued transistor scaling with performance improvement, for future low-power system-on-chip applications.

To mitigate the additional cost due to extra lithography or deposition and etch processes associated with multiple-patterning techniques, **Chapter 4** proposes an innovative cost-efficient patterning method via tilted ion implantation (TII) for achieving sub-lithographic features and/or doubling the density of features, one that is capable of achieving arbitrarily small feature size, self-aligned to pre-existing features on the surface. The patterning resolution limit of TII is investigated via experiments as well as rigorous Monte Carlo process simulations. With feature size below 10 nm experimentally demonstrated and lower line-edge roughness than that of pre-existing masking features on the surface of a substrate, TII-enhanced patterning offers a cost-effective pathway to extend IC technology advancement beyond the 7 nm technology node (sub-40 nm pitch)..

In **Chapter 5**, a survey regarding IC manufacturing cost throughout the semiconductor industry is conducted. Publicly available data and perspectives are collected from representatives of integrated device manufacturers (IDM), foundries, fabless companies, and industry observers. Then two case studies of wafer processing cost are presented: one for FinFET *vs.* iFinFET *vs.* stacked-NW GAA MOSFET, and the other for SADP *vs.* TII double patterning. The results show that the iFinFET technology and TII double patterning technique have significant economic merit.

In **Chapter 6**, the contributions of this dissertation are summarized and suggestons for future work are offered.

1.5 References

[1.1] G. E. Moore, “Cramming more components onto integrated circuits”, *Electronics*, volume 38, number 8, April 19, 1965, pp.114-117.

[1.2] G. E. Moore, “Progress in Digital Integrated Electronics”, Technical Digest 1975. *International Electron Devices Meeting, IEEE*, 1975, pp. 11-13.

[1.3] W. M. Holt, “Moore’s Law: A Path Forward”, *2016 International Solid State Circuits Conference*, paper 1.1.

[1.4] K. Kim, “Silicon Technologies and Solutions for The Data-Driven World”, *2015 International Solid State Circuits Conference*, paper 1.1.

[1.5] C. Auth, C. Allen, A Blattner, D. Bergstrom, M. Brazier, M. Bost, M. Buehler, V. Chikarmane, T. Ghani, T. Glassman, R. Grover, W. Han, D. Hanken, M. Hattendorf, P. Hentges, R. Heussner, J. Hicks, D. Ingerly, P. Jain, S. Jaloviar, R. James, D. Jones, J. Jopling, S. Joshi, C. Kenyon, H. Liu, R. McFadden, B. McIntyre, J. Neirynck, C. Parker, L. Pipes, I. Post, S. Pradhan, M. Prince, S. Ramey, T. Reynolds, J. Roesler, J. Sandford, J. Seiple, P. Smith, C. Thomas, D. Towner, T. Troeger, C. Weber, P. Yashar, K. Zawadzki, and K. Mistry, “A 22nm high performance and low-power CMOS technology featuring fully-depleted tri-gate transistors, self-aligned contacts and high density MIM capacitors,” *VLSI Technol. (VLSIT), 2012 Symp.*, pp. 131–132, 2012.

[1.6] S. Natarajan, M. Agostinelli, S. Akbar, M. Bost, A. Bowonder, V. Chikarmane, S. Chouksey, A. Dasgupta, K. Fischer, Q. Fu, T. Ghani, M. Giles, S. Govindaraju, R. Grover, W. Han, D. Hanken, E. Haralson, M. Haran, M. Heckscher, R. Heussner, P. Jain, R. James, R. Jhaveri, I. Jin, H. Kam, E. Karl, C. Kenyon, M. Liu, Y. Luo, R. Mehandru, S. Morarka, L. Neiberg, P. Packan, A. Paliwal, C. Parker, P. Patel, R. Patel, C. Pelto, L. Pipes, P. Plekhanov, M. Prince, S. Rajamani, J. Sandford, B. Sell, S. Sivakumar, P. Smith, B. Song, K. Tone, T. Troeger, J. Wiedemer, M. Yang, and K. Zhang, “A 14 nm logic technology featuring 2nd-generation FinFET, airgapped interconnects, self-aligned double patterning and a 0.0588 μm^2 SRAM cell size,” in *IEDM Tech. Dig.*, Dec. 2014, pp. 71-73.

[1.7] C. Lin, B. Greene, S. Narasimha, J. Cai, A. Bryant, C. Radens, V. Narayanan, B. Linder, H. Ho, A. Aiyar, E. Alptekin, J. An, M. Aquilino, and R. Bao, V. Basker, N. Breil, M. Brodsky, W. Chang, L. Clevenger, D. Chidambarrao, C. Christiansen, D. Conklin, C. DeWan, H. Dong, L. Economikos, B. Engel, S. Fang, D. Ferrer, A. Friedman, A. Gabor, F. Guarin, X. Guan, M. Hasanuzzaman, J. Hong, D. Hoyos, B. Jagannathan, S. Jain, S-J. Jeng, J. Johnson, B. Kannan, Y. Ke, B. Khan, B. Kim, S. Koswatta, A. Kumar, T. Kwon, U. Kwon, L. Lanzerotti, H-K Lee, W-H. Lee, A. Levesque, W. Li, Z. Li, W. Liu, S. Mahajan, K. McStay, H. Nayfeh, W. Nicoll, G. Northrop, A. Ogino, C. Pei, S. Polvino, R. Ramachandran, Z. Ren, R. Robison, I. Saraf, V. Sardesai, S. Saudari, D. Schepis, C. Sheraw, S. Siddiqui, L. Song, K. Stein, C. Tran, H. Utomo, R. Vega, G. Wang, H. Wang, W. Wang, X. Wang, D. Wehelle-Gamage, E. Woodard, Y. Xu, Y. Yang, N. Zhan, K. Zhao, C. Zhu, K. Boyd, E. Engbrecht, K. Henson, E. Kaste, S. Krishnan, E. Maciejewski, H. Shang, N. Zamdmer, R. Divakaruni, J. Rice, S. Stiffler, P. Agnello, “High Performance 14nm SOI FinFET CMOS Technology with 0.0174μm^2 embedded DRAM and 15 Levels of Cu Metallization,” in *IEDM Tech. Dig*, 2014, pp. 74–76.

[1.8] S. Wu, C. Y. Lin, M. C. Chiang, J. J. Liaw, J. Y. Cheng, S. H. Yang, S. Z. Chang, M. Liang, T. Miyashita, C. H. Tsai, and C. H. Chang, V.S. Chang, Y.K. Wu, J.H. Chen, H.F. Chen, S.Y. Chang, K.H. Pan, R.F. Tsui, C.H. Yao, K.C. Ting, T. Yamamoto, H.T. Huang, T.L. Lee, C.H. Lee, W. Chang, H.M. Lee, C.C. Chen, T. Chang, R. Chen, Y.H. Chiu, M.H. Tsai, S. M. Jang, K.S. Chen, Y. Ku, "An Enhanced 16nm CMOS Technology Featuring 2nd Generation FinFET Transistors and Advanced Cu/ low-k Interconnect for Low Power and High Performance Applications," in *IEDM Tech. Dig*, 2014, pp. 48–51.

[1.9] K. Mistry, "Tri-Gate Transistors: Enabling Moore's Law at 22nm and Beyond", *SEMICON West*, July 2014.

[1.10] "International Technology Roadmap for Semiconductors (ITRS), 2013." [Online]. Available: http://public.itrs.net/.

[1.11] C. Dupré, A. Hubert, S. Becu, M. Jublot, V. Maffini-Alvaro, C. Vizioz, F. Aussenac, C. Arvet, S. Barnola, J.-M. Hartmann, G. Garnier, F. Allain, J.-P. Colonna, M. Rivoire, L. Baud, S. Pauliac, V. Loup, T. Chevolleau, P. Rivallin, B. Guillaumot, G. Ghibaudo, O. Faynot, T. Ernst, and S. Deleonibus, "15nm-diameter 3D Stacked Nanowires with Independent Gates Operation: ΦFET," in *IEDM Tech. Dig*, 2008, pp. 1 – 4.

[1.12] C. Wagner and N. Harned, "EUV Lithography: Lithography Gets Extreme," *Nature Photonics 2010*, 4, pp. 24-26.

[1.13] Y.-K., Choi, T.-J., King, C. Hu, "A Spacer Patterning Technology for Nanoscale CMOS", *IEEE Trans. Electron Devices* 2002, 49(3), pp. 436-441.

[1.14] C. Bencher, Y. Chen, H. Dai, W. Montgomery, L. Huli, "22nm Half-Pitch Patterning by CVD Spacer Self Alignment Double Patterning (SADP)". *Proc. SPIE 6924, Optical Microlithography XXI* 2008, 6924, 69244E1-69244E7.

[1.15] B. Turkot, S. Carson, A. Lio, T. Liang, M. Phillips, B. McCool, E. Stenehjem, T. Crimmins, G. Zhang, G. Sivakumar, "EUV Progress Toward HVM Readiness", *Proc. SPIE 9776, Extreme Ultraviolet (EUV) Lithography VII* 2016, 9776, 977602-01-977602-09.

Chapter 2

10 nm L_g Multi-Gate MOSFET Technologies

2.1 Introduction

As Tri-gate MOSFETs (also known as FinFETs [2.1]) already have been adopted at the 22 nm CMOS technology node [2.2], it is generally accepted that multi-gate structures will be necessary to scale the MOSFET gate length down to 10 nm and below. In this regime, the GAA MOSFET design can achieve superior electrostatic integrity [2.3]. It is well known that process-induced variation in transistor performance is one of the significant challenges to further transistor miniaturization [2.4]. This chapter compares the performance and variability of a GAA MOSFET against that of an ideal silicon-on-insulator (SOI) FinFET design at 10 nm gate length (L_g), accounting for systematic and random variations. The benefits of GAA MOSFET technology for lowering the minimum operating voltage (V_{min}) and area of a six-transistor (6-T) SRAM cell to facilitate increased transistor density following Moore's Law are assessed.

2.2 GAA and SOI FinFET Design Optimization

2.2.1 Nominal MOSFET Design

Fig.2.1 schematically illustrates the SOI FinFET and GAA MOSFET structures which were studied via technology computer-aided design (TCAD) three-dimensional (3-D) device simulation [2.5] in this work. The gate length (L_g) is 10 nm, and equivalent oxide thickness (EOT) is 0.62 nm, based on ITRS specifications for the 4/3 nm technology node [2.6]. Raised-source/drain regions are assumed to be formed by selective epitaxial growth *in-situ* doped Si (2×10^{20} cm^{-3}) [2.7]; the source/drain extensions have a Gaussian lateral (1-D) doping profile with peak concentration at the edge of the raised-source/drain regions. Ohmic contacts (4×10^{-9} Ω-cm^2) are made to the top surfaces of the raised-source/drain regions. The gate pitch is fixed at 30 nm. The nominal supply voltage V_{DD} is 0.68 V.

The concept of an electrostatic scale length (λ) was proposed for both tri-gate [2.8-2.9] and GAA [2.10-2.11] MOSFETs. To make a fair comparison, the scale length is selected to be the same (λ = 3.3 nm) for both FinFETs and GAA MOSFETs in this work. A previous study showed that a wide and short channel design is advantageous for the GAA MOSFET [2.12]. For SOI FinFETs, the fin aspect ratio (H_{si}/W_{si}) is selected to be ~2.5 [2.13]. **Fig. 2.2(a)** delineates the channel dimensions required to achieve λ = 3.3 nm

scale length for FinFETs and GAA MOSFETs. Channel region cross-sections and dimensions of nominal FinFET and GAA MOSFET are shown in **Fig. 2.2(b)**.

Device simulations are performed with Sentaurus using drift-diffusion transport, a density gradient model for quantum confinement, bandgap narrowing effect, and Philips and high-field degradation models for mobility [2.5]. Since strain-induced mobility enhancements diminish with L_g, they are not included. Simulated device performance parameters are summarized in **Table 2.1**. Transistor threshold voltage V_T is defined as the voltage where drain current I_D is 100 nA × W_{eff}/L_g. Transistor ON-state current I_{ON} is defined as drain current I_D, for gate voltage $V_G = V_{DD}$ and drain voltage $V_D = V_{DD}$. Transistor OFF-state current I_{OFF} is defined as drain current I_D, for gate voltage $V_G = 0$ V and drain voltage $V_D = V_{DD}$. Drain-induced barrier lowering (DIBL) and sub-threshold swing (SS) are also included.

Fig. 2.1 Schematic views of the SOI-FinFET (left) and GAA-MOSFET (right) structures investigated in this work.

(a) (b)

Fig. 2.2 (a) Channel region dimensions required to achieve 3.3 nm scale length for FinFETs (red) and GAA MOSFETs (blue). Stars indicate the dimensions used in this work. (b) Channel region cross-sections of the SOI-FinFET (left, W_{Si}=6.5 nm, H_{Si}=16 nm) & GAA (right, W_{Si}=16 nm, H_{Si}=7.5 nm) MOSFET.

Table 2.1 DEVICE PERFORMANCE COMPARISON

	SOI FINFET		GAA MOSFET	
	NMOS	PMOS	NMOS	PMOS
V_T (V)	0.17	-0.18	0.17	-0.18
I_{OFF} (nA)	3.8	-3.8	4.7	-4.7
I_{ON} (μA)	22.4	-15.8	28.3	-19.2
SS (mV/dec)	83.4	86.7	84.3	87.8
DIBL (mV/V)	73.0	73.0	69.8	71.4

2.2.2 Compact Model Calibration

A physically based analytical model [2.14] is calibrated to the simulated I-V characteristics for both the linear and saturation regions of operation, and to predict 6-T SRAM cell performance and to estimate cell yield [2.15]. Fig. **2.3** shows that this calibrated transistor I-V model well matches the TCAD simulations for both FinFETs and GAA MOSFETs.

Fig. 2.3 Comparison of the calibrated analytical transistor I-V model against 3-D TCAD simulations for FinFETs (left) and GAA MOSFETs (right). A good match is seen between the compact model and 3-D TCAD device simulations.

2.3 Impact of Process-Induced Variations

2.3.1 Impact of Systematic Variations

Fig. 2.4 shows L_g dependence of threshold voltage (V_T, defined as the voltage, where drain current I_D is 100 nA × W_{eff}/L_g) and off current (I_{OFF}, defined as drain current I_D, for gate voltage V_G = 0 V and drain voltage $V_D = V_{DD}$). The short channel effect is seen to be accurately captured by the compact model for both FinFETs and GAA MOSFETs.

For multi-gate MOSFETs, V_T is dependent on the channel width W_{Si} since the side gates influence the channel potential. Since the FinFET relies on a narrow fin (small W_{Si}) to suppress off-state leakage current, it is more sensitive to W_{Si} variations than the GAA MOSFET, which has a short and wide channel design in this work. **Fig. 2.5** shows V_T and I_{OFF} sensitivity to W_{Si}. Quantum confinement effects are seen in both FinFETs and GAA MOSFETs, especially pronounced when W_{Si} is below 10 nm. Again, a good match is seen between the compact model and the 3-D TCAD device simulations.

Fig. 2.4 (a) V_T roll-off characteristics for SOI-FinFETs (left) and GAA MOSFETs (right). (b) I_{OFF} sensitivity to L_g variation for SOI-FinFETs (left) and GAA MOSFETs (right).

Fig. 2.5 (a) Impact of variations in fin/stripe width for V_T of SOI-FinFETs (left) and GAA MOSFETs (right). (b) I_{OFF} sensitivity to W_{Si} variation for SOI-FinFETs (left) and GAA MOSFETs (right).

2.3.2 Impact of Random Variations

In recent years, the importance of suppressing random performance variations has increased because they ultimately limit the extent to which the supply voltage (hence power consumption) can be reduced. The minimum operating voltage (V_{min}) for memory elements such as SRAM cells is set by manufacturing yield requirements [2.4]. Among all random variation sources, gate line-edge-roughness (LER) [2.16-2.17], random dopant fluctuation (RDF) [2.18-2.20] and metal gate work function variation (WFV) [2.21-2.25] usually have the greatest impact. Therefore, these sources of variation are investigated herein for both FinFETs and GAA MOSFETs.

To study the impact of gate line-edge-roughness (G-LER), 3-D device simulations with 1 nm (root mean square value) roughness and 10 nm correlation length (following ITRS specifications for lithography [2.6]) are performed for 250 devices, following the methodology described in [2.16]. The G-LER induced V_T variation (σV_T) is shown in **Fig. 2.6(a)** for FinFETs and GAA MOSFETs.

The impedance field method (IFM) [2.5], [2.20] is used to assess the impact of RDF for 5000 devices of each design. The results shown in **Fig. 2.6(b)** show that RDF-induced variation is relatively small. This is because the channel regions of the FinFET and GAA MOSFET devices have low nominal dopant concentration (1×10^{15} cm^{-3}) in this study.

Previous studies have shown theoretically [2.21-2.24] and experimentally [2.25] that WFV will be the dominant source of random variation in nanometer-scale devices, especially with undoped channel regions as for the devices in this study. In this work, the gate material is assumed to be TiN. The work function value and probability of occurrence for each grain orientation is summarized in [2.24]. The impedance field method (IFM) [2.5], [2.26] is used again to assess WFV induced variations (**Fig. 6(c)**) for 5000 devices of each design.

Fig. 2.6 Simulated transfer characteristics in the saturation region of operation showing the impact of random variability sources (a) G-LER, (b) RDF, (c) WFV for n-channel SOI-FinFET (left) and GAA-MOSFET (right). The black line represents the I-V characteristic for the nominal design. The red lines represent the first 200 simulated variation cases.

Table 2.2 shows in detail the standard deviations of V_T (in the saturation region of operation), I_{ON} and log (I_{OFF}) due to each random variation source for each device structure (SOI-FinFET and GAA-MOSFET) and both transistor types (NMOS and PMOS).

Table 2.2 RANDOM VARIATION SOURCE COMPARISON

SOI-FinFET	NMOS			PMOS		
	G-LER	RDF	WFV	G-LER	RDF	WFV
σV_T (mV)	2.7	5.7	24	2.6	5.9	24
σI_{ON} (A)	3.29×10^{-7}	5.79×10^{-7}	1.38×10^{-6}	1.78×10^{-7}	8.57×10^{-7}	8.93×10^{-7}
$\sigma \log(I_{OFF})$ (A)	5.46×10^{-2}	1.03×10^{-1}	3.02×10^{-1}	5.41×10^{-2}	9.82×10^{-2}	2.89×10^{-1}
GAA-MOSFET	**NMOS**			**PMOS**		
	G-LER	RDF	WFV	G-LER	RDF	WFV
σV_T (mV)	2.4	5.5	22	2.4	5.7	22
σI_{ON} (A)	4.49×10^{-7}	5.11×10^{-7}	1.67×10^{-6}	2.21×10^{-7}	5.33×10^{-7}	1.05×10^{-6}
$\sigma \log(I_{OFF})$ (A)	4.83×10^{-2}	9.03×10^{-2}	2.76×10^{-1}	4.79×10^{-2}	1.03×10^{-1}	2.65×10^{-1}

2.4 6-T SRAM Cell Performance and Yield

A 6-T SRAM cell comprises 2 pull-up (M_{PU}), 2 pull-down (M_{PD}) and 2 access/pass-gate (M_{PG}) transistors, as shown in **Fig. 2.7**. In this work, the read static noise margin (SNM) and write-ability current (I_w) of 6-T SRAM cells implemented with 10 nm L_g FinFETs and GAA MOSFETs are investigated using TCAD mixed-mode simulations [2.5], and also using the aforementioned calibrated compact model [2.14].

Fig. 2.7 Circuit diagram of a 6-T SRAM cell

2.4.1 FinFET 6-T SRAM Cell Design

Local adjustments in channel width (fin height) to separately tune the drive currents of individual transistors within a 6-T SRAM cell are not as straightforward to implement in FinFET technology *vs.* a (quasi) planar MOSFET technology. Neither is FinFET V_T adjustment via doping an attractive option, because it requires a dopant concentration much greater than 10^{18} cm^{-3} (as shown in **Fig. 2.8**) which would result in large RDF and degraded performance due to lower carrier effective mobilities. V_T adjustment via gate work function tuning would require different gate materials and hence increased process complexity. In this work, the beta ratio of FinFET-based 6-T SRAM cells is practically tuned as is done in [2.27], by adjusting the number of fins in the pull-down (PD) devices.

Fig. 2.8 Doping dependence of threshold voltage and saturation current for an n-channel SOI-FinFET.

Fig. 2.9 (a) shows how the static noise margin (SNM) and write-ability current (I_w) change with the number of fins in the PD devices. The values of SNM and I_w are summarized in **Table 2.3**. The 2-fin PD device design provides for a better trade-off between read stability and write-ability, hence for lower voltage operation.

Fig. 2.9 Butterfly plots (left) and write-N curves (right) for 6-T SRAM cells implemented with SOI FinFETs (a) or GAA MOSFETs (b). Good agreement in SNM and I_W between TCAD (mixed-mode) simulations and the compact model is seen.

TABLE 2.3 SUMMARY COMPARISON OF 6-T SRAM CELL DESIGNS WITH FINFETS AND GAA MOSFETS

	FinFET SRAM		GAA SRAM
	1 PD	**2 PD**	**W_{PD}=20nm**
SNM (mV)	**101.8**	**142.3**	**140.9**
I_W (μA)	**15.5**	**12.4**	**11.8**

2.4.2 GAA MOSFET 6-T SRAM Cell Design

The width of a GAA MOSFET can be adjusted to tune its drive current, if a short and wide channel design is employed [2.12], [2.28]. In this work, the width of the PD devices in the GAA-MOSFET-based 6-T SRAM cell is adjusted to achieve a good trade-off between read stability and write-ability. The resultant butterfly plot (for SNM) and “write-N” curve (for I_w) are shown in **Fig. 2.9 (b)**.

2.4.3 6-T SRAM Cell Area Comparison

The 6-T SRAM cell layouts described in [2.29], [2.30] are scaled herein to be appropriately sized for 10 nm gate length devices [2.6]. **Fig. 2.10** compares the half-cell layouts for (2-fin PD) FinFET and GAA-MOSFET-based 6-T SRAM. Due to fin pitch (30 nm) limitations, the FinFET-based cell occupies ~20% more area.

Fig. 2.10. Half-cell layouts for 6-T SRAM implemented with FinFETs (left) or GAA MOSFETs (right).

2.4.4 6-T SRAM Cell Yield Comparison

To estimate 6-T SRAM cell read and write yield, the compact model is employed to calculate the cell sigma, defined as the minimum number of standard deviations (for any combination of variation sources) that can result in a read failure or a write failure, accounting for process-induced variations in device width and gate length (assuming Gaussian distributions with 3σ = 10% of nominal value) as well as random variations in V_T due to G-LER, RDF and WFV. V_{min}, defined as the lowest value of V_{DD} that meets the six-sigma yield requirement for both read and write operations in large capacity SRAM, is found to be 0.52 V for GAA-MOSFET-based SRAM and 0.6 V for FinFET-based SRAM (**Fig. 2.11**). Further optimization of the trade-off between read and write yield (by changing the PD device width to 16 nm) can reduce V_{min} to ~0.51 V for the GAA-MOSFET-based SRAM cell by better balanced read and write yields.

Fig. 2.11 Read yield (squares) and write yield (triangles) of 6-T SRAM cells implemented with FinFETs (blue) or GAA MOSFETs (red, green) vs. cell operating voltage.

2.5 Summary

A variation-aware comparison between GAA-MOSFET and SOI-FinFET technologies at 10 nm L_g reveals that GAA-MOSFETs should provide for ~0.1 V reduction in 6-T SRAM cell operating voltage, with ~20% reduction in cell area.

2.6 References

[2.1] D. Hisamoto, W.-C. Lee, J. Kedzierski, H. Takeuchi, K. Asano, C. Kuo, E. Anderson, T.-J. King, J. Bokor, and C. Hu, "FinFET— A self-aligned double-gate MOSFET scalable to 20 nm," *IEEE Trans. Electron Devices*, vol. 47, no. 12, pp. 2320–2325, Dec. 2000.

[2.2] C. Auth, C. Allen, a Blattner, D. Bergstrom, M. Brazier, M. Bost, M. Buehler, V. Chikarmane, T. Ghani, T. Glassman, R. Grover, W. Han, D. Hanken, M. Hattendorf, P. Hentges, R. Heussner, J. Hicks, D. Ingerly, P. Jain, S. Jaloviar, R. James, D. Jones, J. Jopling, S. Joshi, C. Kenyon, H. Liu, R. McFadden, B. McIntyre, J. Neirynck, C. Parker, L. Pipes, I. Post, S. Pradhan, M. Prince, S. Ramey, T. Reynolds, J. Roesler, J. Sandford, J. Seiple, P. Smith, C. Thomas, D. Towner, T. Troeger, C. Weber, P. Yashar, K. Zawadzki, and K. Mistry, "A 22nm high performance and low-power CMOS technology featuring fully-depleted tri-gate transistors, self-aligned contacts and high density MIM capacitors," *VLSI Technol. (VLSIT), 2012 Symp.*, pp. 131–132, 2012.

[2.3] M. Li, K. H. Yeo, S. D. Suk, Y.-Y Yeoh, D.-W. Kim, T. Y. Chung, K. –S. Oh, and W.-S. Lee, “Sub-10nm gate-all-around CMOS nanowire transistors on bulk Si substrate,” *VLSI Technol. (VLSIT), 2009 Symp.*, pp. 94-95, 2009.
[2.4] K. Kuhn, M. D. Giles, D. Becher, P. Kolar, A. Kornfeld, R. Kotlyar, S. T. Ma, A. Maheshwari, and S. Mudanai, “Process technology variation,” *IEEE Trans. Electron Devices*, vol. 58, no. 8, pp. 2197–2208, Aug. 2011.
[2.5] Sentaurus User’s Manual, Synopsys, Inc., 2013.
[2.6] International Technology Roadmap for Semiconductors (ITRS), 2011 (http://public.itrs.net/)
[2.7] A. Hokazono, K. Ohuchi, K. Miyano, I. Mizushima, Y. Tsunashima, and Y. Toyoshima, “Source/drain engineering for sub-100nm CMOS using selective epitaxial growth technique,” in *IEDM Tech. Dig.*, pp. 243-246, 2000.
[2.8] J. Colinge, “Multiple-gate SOI MOSFETs,” *Solid-State Electrons*, vol. 48, pp. 897–905, 2004.
[2.9] M.-D. Ko, C.-W. Sohn, C.-K. Baek, and Y.-H. Jeong, “Study on a Scaling Length Model for Tapered Tri-Gate FinFET Based on 3-D Simulation and Analytical Analysis,” *IEEE Trans. Electron Devices*, vol. 60, no. 9, pp. 2721–2727, Sep. 2013.
[2.10] C. P. Auth, and J. D. Plummer, “Scaling theory for cylindrical, fully-depleted, surrounding-gate MOSFET's,” *IEEE Electron Device Letters*, vol. 18, pp. 74-76, 1997.
[2.11] S. Bangsaruntip, G. M. Cohen, A. Majumdar, and J. W. Sleight, “Universality of Short-Channel Effects in Undoped-Body Silicon Nanowire MOSFETs,” *IEEE Electron Device Lett.*, vol. 31, no. 9, pp. 903–905, Sep. 2010.
[2.12] Y.-B. Liao, M.-H. Chiang, N. Damrongplasit, T.-J. King Liu, and W.-C. Hsu, “6-T SRAM cell design with gate-all-around silicon nanowire MOSFETs,” in *VLSI-TSA Symp.*, pp. 1-2, 2013.
[2.13] V. S. Basker, T. Standaert, H. Kawasaki, C.-C. Yeh, K. Maitra, T. Yamashita, J. Faltermeier, H. Adhikari, H. Jagannathan, J. Wang, H. Sunamura, S. Kanakasabapathy, S. Schmitz, J. Cummings, A. Inada, C.-H. Lin, P. Kulkarni, Y. Zhu, J. Kuss, T. Yamamoto, A. Kumar, J. Wahl, A. Yagishita, L. F. Edge, R. H. Kim, E. Mclellan, S. J. Holmes, R. C. Johnson, T. Levin, J. Demarest, M. Hane, M. Takayanagi, M. Colburn, V. K. Paruchuri, R. J. Miller, H. Bu, B. Doris, D. McHerron, E. Leobandung, and J. O'Neill, “A 0.063 um^2 FinFET SRAM cell demonstration with conventional lithography using a novel integration scheme with aggressively scaled fin and gate pitch,” in *Proc. Symp. VLSI Technol.,* pp.19 -20 2010.
[2.14] A. E. Carlson, “Device and circuit techniques for reducing variation in nanoscale SRAM,” Ph. D. dissertation, *Dept. EECS Univ. California Berkeley*, Berkeley, CA, 2008.
[2.15] C. Shin, “Advanced MOSFET designs and implications for SRAM scaling,” Ph. D. dissertation, *Dep. EECS Univ. California Berkeley*, Berkeley, CA, 2011.
[2.16] A. Asenov, S. Kaya, and A. R. Brown, “Intrinsic parameter fluctuations in decananometer MOSFETs introduced by gate line edge roughness,” *IEEE Trans. Electron Devices*, vol. 50, pp.1254-1260, 2003.

[2.17] K. Patel, T.-J. K. Liu, C. J. Spanos, "Gate line edge roughness model for estimation of FinFET performance variability," *IEEE Transactions on Electron Devices*, Vol. 56, No. 12, pp. 3055-3063, 2009.
[2.18] H.-S. Wong and Y. Taur, "Three-dimensional "atomistic" simulation of discrete random dopant distribution effects in sub-0.1 ìm MOSFETs," in *IEDM Tech. Dig*., pp. 705–708. 1993.
[2.19] A. Asenov, "Random dopant induced threshold voltage lowering and fluctuations in sub-0.1-μm MOSFETs: A 3-D 'atomistic simulation study'," *IEEE Trans. Electron Devices*, vol. 45, no. 12, pp. 2505–2513, Dec. 1998.
[2.20] A. Wettstein, O. Penzin, E. Lyumkis, and W. Fichtner, "Random dopant fluctuation modelling with the impedance field method," in *Proc. Simul. Semicond. Process. Devices (SISPAD)*, Sep. 2003, pp. 91–94.
[2.21] H. Dadgour, K. Endo, V. De, and K. Banerjee, "Modeling and analysis of grain-orientation effects in emerging metal-gate devices and implications for SRAM reliability," in *Proc. IEEE Int. Electron Dev. Meeting*, Dec. 2008, pp. 1–4.
[2.22] H. Dadgour, K. Endo, V. De, and K. Banerjee, "Grain-orientation induced work function variation in nanoscale metal-gate transistors part I: Modeling, analysis, and experimental validation," *IEEE Trans. Electron Dev.*, vol. 57, no. 10, pp. 2504–2514, Oct. 2010.
[2.23] X. Wang, A. R. Brown, N. Idris, S. Markov, G. Roy, and A. Asenov, "Statistical threshold-voltage variability in scaled decananometer bulk HKMG MOSFETs: A full-scale 3-D simulation scaling study," *IEEE Trans. Electron Dev.*, vol. 58, no. 8, pp. 2293–2301, Aug. 2011.
[2.24] X. Zhang, J. Li, M. Grubbs, M. Deal, B. Magyari-Köpe, B. M. Clemens, and Y. Nishi, "Physical model of the impact of metal grain work function variability on emerging dual metal gate MOSFETs and its implication for SRAM reliability," in *IEDM Tech. Dig.*, 2009, pp. 57–60.
[2.25] T. Matsukawa, S. O'uchi, K. Endo, Y. Ishikawa, H. Yamauchi, Y.-X. Liu, J. Tsukada, K. Sakamoto, and M. Masahara, "Comprehensive analysis of variability sources of FinFET characteristics," in *Proc. Symp. VLSI Technol.*, Jun. 2009, pp. 118–119.
[2.26] K. El Sayed, E. Lyumkis, and A. Wettstein, "Modeling Statistical Variability with the Impedance Field Method," in *International Conference on Simulation of Semiconductor Processes and Devices (SISPAD)*, Denver, CO, USA, pp. 205–208, September 2012.
[2.27] C.-H. Jan, U. Bhattacharya, R. Brain, S.-J. Choi, G. Curello, G. Gupta, W. Hafez, M. Jang, M. Kang, K. Komeyli, T. Leo, N. Nidhi, L. Pan, J. Park, K. Phoa, A. Rahman, C. Staus, H. Tashiro, C. Tsai, P. Vandervoorn, L. Yang, J.-Y. Yeh, and P. Bai, "A 22nm SoC platform technology featuring 3-D tri-gate and high-k/metal gate, optimized for ultra low power, high performance and high density SoC application," in *Proc. Int. Electron. Devices Meeting*, Dec. 2012, pp. 44–47.
[2.28] Y.-B. Liao, M.-H. Chiang, N. Damrongplasit, W.-C. Hsu, and T.-J. King Liu, "Design of Gate-All-Around Silicon MOSFETs for 6-T SRAM Area Efficiency and Yield," *IEEE Trans. Electron Devices*, vol. 61, pp. 2371-2377, 2014.
[2.29] C. Shin, M. H. Cho, Y. Tsukamoto, B.-Y. Nguyen, C. Mazure, B. Nikolic, and T.-J. King Liu, "Performance and area scaling benefits of FD-SOI technology for 6-T

SRAM cells at the 22-nm node," *IEEE Trans. Electron Devices*, vol. 57, pp. 1301-1309, 2010.
[2.30] C. Shin, N. Damrongplasit, X, Sun, Y. Tsukamoto, B. Nikolic, and T.-J. King Liu, "Performance and yield benefits of quasi-planar bulk CMOS technology for 6-T SRAM at the 22-nm node," *IEEE Trans. Electron Devices*, vol. 28, pp. 1846-1854, 2011.

Chapter 3

Inserted-Oxide FinFET (iFinFET)

3.1 Introduction

The FinFET multi-gate transistor structure is widely used in the most advanced (16/14 nm generation) complementary metal-oxide-semiconductor (CMOS) technologies today [3.1-3.3]. The gate-all-around (GAA) field-effect transistor (FET) is anticipated to be adopted in future generations to enable ultimate gate-length scaling due to its superior electrostatic integrity [3.4]. However, to achieve comparable (or even better) layout area efficiency as a FinFET, a GAA FET must comprise multiple stacked nanowires (NWs) [3.5], with significant added fabrication process complexity. This is because stacked NWs are formed by epitaxial growth of relatively thick (>10 nm) $Si_{1-x}Ge_x$ sacrificial layers between Si channel layers to accommodate gate-dielectric/gate-metal/gate-dielectric layers in-between the NWs, so that fin structures with very high aspect ratio must be etched prior to selective removal of the sacrificial $Si_{1-x}Ge_x$ layers. Also, it will be more difficult to implement multiple gate-oxide thicknesses as needed for system-on-chip (SoC) applications, in GAA FET technology. This chapter proposes an evolutionary multi-gate transistor design, the inserted-oxide FinFET (iFinFET) [3.6-3.8], to circumvent these challenges by providing for improved gate control without any added fabrication process complexity. The performance characteristics of the iFinFET are benchmarked against the FinFET and also the stacked-NW GAA FET, for both n-channel (NMOS) and p-channel (PMOS) transistors, via technology computer-aided design (TCAD) 3-D device simulations using Sentaurus Device [3.9]. The results show that the iFinFET provides for improved electrostatic integrity relative to the FinFET, but with substantially less gate capacitance penalty relative to the GAA FET. Thus, iFinFET technology offers a technological pathway for continued transistor scaling with performance improvement, for future low-power SoC applications. The effects of process-induced variations on iFinFET performance also are investigated.

3.2 iFinFET Structure and Fabrication

Fig. 3.1 shows the cross-sections across the channel region for a conventional bulk FinFET, an iFinFET and a stacked-NW GAA FET. The effective channel width (W_{eff}, defined as the Si outer perimeter above the shallow trench isolation oxide) is the same for all three devices. The iFinFET can be fabricated using a process (**Fig. 3.2**) identical to that for a conventional bulk FinFET but starting with a multi-SOI (silicon-on-insulator on silicon-on-insulator) substrate. Since SiO_2 is usually etched during the standard cleaning processes prior to gate-stack formation, the inserted-oxide layers would be slightly recessed (*e.g.* by 2 nm as illustrated in **Fig. 3.1**) in practice, so that the high-

permittivity (high-k) gate dielectric layer, formed by atomic layer deposition, would penetrate in-between the Si channels as shown in **Fig. 3.1**. The inserted dielectric layers in the iFinFET allow fringing electric fields to provide for improved gate control, and can be much thinner than the inserted dielectric/metal/dielectric layers in a GAA FET [3.6].

Fig. 3.1 Cross-sectional views across the fin channel region of multi-gate transistors.

Fig. 3.2 Schematics illustrating key process steps in the fabrication flow of a conventional bulk FinFET (after [3.9]). The iFinFET can be fabricated using an identical process but starting with a multi-SOI (silicon-on-insulator on silicon-on-insulator) substrate.

Fig. 3.3 shows the simulated device cross-sections along the channel direction. The heavily doped source and drain (S/D) regions are assumed to be epitaxially regrown and *in-situ*-doped; therefore, the S/D regions of the iFinFET do not comprise inserted-oxide layers. (Since epitaxy occurs both vertically and laterally, no significant issues are foreseen to be caused by the presence of the relatively thin (3 nm) inserted-oxide layers.) For the stacked-NW GAA FET, the metal gate thickness in-between the Si NWs is assumed to be 6 nm. The gate length (L_g) is 12 nm, appropriate for the 4/3 nm technology node [3.4]. The gate dielectric comprises a 0.5 nm SiO_2 layer and a 1.28 nm

HfO_2 (relative permittivity = 25) layer, and thus has an equivalent SiO_2 thickness of 0.7 nm. The metal gate work function is adjusted to achieve off-state leakage current I_{OFF} = 20 pA/µm. Each of the devices is assumed to be fabricated on a standard Si {100} wafer, such that the top Si surface is along a {100} plane and the sidewall Si surfaces are along {110} planes. The relative dielectric permittivity of the gate-sidewall spacer material is assumed to be 5 [3.10-3.11]. The n-channel FETs comprise Si channel and S/D regions, whereas the p-channel FETs comprise Si channel and $Si_{0.5}Ge_{0.5}$ S/D regions. The average stress in the channel region along the direction of the current flow is 2 GPa (tensile for NMOSFETs, compressive for PMOSFETs). The nominal supply voltage V_{DD} is 0.75 V, based on the roadmap for low-power CMOS technology [3.4]. Key device design parameter values are listed in **Table 3.1**. To provide for a fair comparison, the effective channel length (L_{eff}, defined as the distance between the locations where the S/D doping falls to 2×10^{19} cm^{-3}) is selected to optimize the tradeoff between good short-channel control and low series resistance.

Fig. 3.3 Cross-sectional views along the channel direction of multi-gate transistors. The net dopant concentration is represented in color using a hyperbolic arcsine (asinh) scale. The black dashed line indicates the depth corresponding to the surface of the shallow trench isolation oxide.

TABLE 3.1 TRANSISTOR DESIGN PARAMETER VALUES

Device Parameter	FinFET / iFinFET / GAA
Nominal Gate Length, L_g (nm)	12
Fin/NW Width, W_{fin} (nm)	6
Total Si Height above STI (nm)	18
Gate Pitch (nm)	35
Equivalent Oxide Thickness (nm)	0.7
Inserted-Oxide Thickness (nm)	0 / 3 / 0
Metal Thickness between NWs (nm)	0 / 0 / 6
Specific Contact Resistivity (Ω-cm^2)	3.5×10^{-9}

3.3 Comparison of FinFET, iFinFET and GAA MOSFET Performance Characteristics

The performance characteristics of the iFinFET are benchmarked against the FinFET and also the stacked-NW GAA FET, for both n-channel (NMOS) and p-channel (PMOS) transistors, via technology computer-aided design (TCAD) 3-D device simulations using Sentaurus Device [3.9], with drift-diffusion carrier transport parameters calibrated against Monte Carlo simulations to model ballistic transport, band-to-band tunneling parameters calibrated against non-equilibrium Green's Function simulations [3.12] and the density gradient electrostatic quantum correction parameters calibrated against empirical pseudopotential simulations [3.13]. The density gradient method has been shown to provide a computationally efficient approximation to more rigorous Schrödinger-Poisson solutions to 1-, 2-, and 3-dimensional silicon structures [3.14].

3.3.1 Why Inserted-Oxide, not Inserted-HK Dielectric?

Fig. 3.4 shows the electrostatic potential contours and the electric field lines distribution for each of the multi-gate transistor designs. It is clear that fringing electric fields enhance the coupling between the gate and channel regions in the iFinFET.

Fig. 3.4 Comparison of electrostatic potential distributions (represented in color) and electric field lines (arrows) in the linear region of operation (V_{GS} = 0.75 V, V_{DS} = 0.05 V). The electric potential in the gate is not shown for simplicity.

Since the performance advantages of the iFinFET stem from increased capacitive coupling between the gate and channel regions, it is worthwhile to consider high-permittivity (high-k) materials such as Si_3N_4 and HfO_2 for the inserted dielectric layers. **Fig. 3.5** shows how drain-induced barrier lowering (DIBL) and sub-threshold swing (SS) depend on the material and thickness of the inserted-dielectric layers. A higher-k material results in worse electrostatic integrity because it results in higher capacitive coupling between the drain and channel regions. SiO_2 as the inserted dielectric material provides for the highest on-state drive current I_{ON} (**Fig. 3.6(a)**) and lowest intrinsic delay (**Fig. 3.6(b)**). Together with the high quality of the Si/SiO_2 interface in practice, these results indicate that SiO_2 is the optimal inserted-dielectric material for the iFinFET.

Fig. 3.5 Impact of the inserted-dielectric material and thickness on drain-induced barrier lowering (DIBL) and sub-threshold swing (SS) for n-channel iFinFETs.

Fig. 3.6 Impact of the inserted-dielectric material and thickness on (a) on-state drive current (I_{ON}) and (b) intrinsic delay for n-channel iFinFETs.

3.3.2 Comparison of Transistor Performance Characteristics

Simulated transfer (I_D-V_{GS}) and output (I_D-V_{DS}) characteristics are shown in **Fig. 3.7** and **Fig. 3.8**, respectively. Improved gate control results in steeper subthreshold swing and less DIBL. Therefore, the iFinFETs have substantially larger ON-state drive current (I_{ON}) than the FinFETs, by ~20% for NMOS and ~14% for PMOS. Larger improvement for NMOS is seen because the fringing electric fields through the inserted-oxide layers increase capacitive coupling to Si {100} surfaces, which have higher electron mobility relative to the sidewall {110} surfaces, but lower hole mobility [3.15]. Stacked-NW GAA FETs have even larger I_{ON}, due to superior gate control of the electric potential in the channel regions. **Fig. 3.9** compares the tradeoff between high I_{ON} and low I_{OFF} for the various multi-gate FET designs.

Fig. 3.7 Simulated transistor transfer characteristics, for V_{DS} = 0.05 V (dashed lines) and V_{DS} = 0.75 V (solid lines). Current is normalized to W_{eff}. The gate work function is adjusted to achieve I_{OFF} = 20 pA/μm.

Fig. 3.8 Simulated transistor output characteristics. Current is normalized to W_{eff}. The gate work function is adjusted to achieve I_{OFF} = 20 pA/μm.

Fig. 3.9 Comparison of (a) PMOS and (b) NMOS I_{OFF} *vs.* I_{ON} for multi-gate FET technologies. (The data re obtained by varying L_g.)

The larger I_{ON} of the stacked-NW GAA comes at the cost of higher gate capacitance, as shown in **Fig. 3.10**. Of all the multi-gate MOSFETs, the stacked-NW GAA FET has the largest total gate capacitance (C_{gg}) and the largest gate-to-drain "Miller" capacitance (C_{gd}); as a result, it has the largest intrinsic delay, ~9% and ~17% larger than the FinFET for NMOS and PMOS, respectively (**Fig. 3.11**). It should be noted that this issue for the stacked-NW GAA FET would be exacerbated if the metal gate layers in-between the NWs are thicker. On the other hand, although C_{gg} is also larger for the iFinFET than for the FinFET, the larger I_{ON} of the iFinFET compensates well for this, so that there is negligible intrinsic delay penalty (0% for NMOS and ~4% for PMOS).

Fig. 3.10 Comparison of total gate capacitance (C_{gg}) and gate-to-drain capacitance (C_{gd}) for multi-gate FET technologies.

Fig. 3.11 Comparison of transistor intrinsic delay ($C_{gg}V_{DD}/I_{ON}$) for multi-gate FET technologies.

Transistor performance parameters pertinent to analog/mixed-signal applications are also examined herein since SoC products are becoming ubiquitous. Due to improved electrostatic integrity, the iFinFET and the stacked-NW GAA FET provide for improvements in intrinsic gain (g_m/g_{ds}) over the FinFET, by ~15% and ~36% for NMOS and by ~7% and ~27% for PMOS, respectively (**Fig. 3.12**).

Fig. 3.12 Comparison of transistor intrinsic gain (g_m/g_{ds}) for multi-gate FET technologies.

Table 3.2 provides a summary comparison of key transistor performance parameters.

TABLE 3.2 COMPARISON OF TRANSISTOR PERFORMANCE PARAMETERS

	FinFET		iFinFET		GAA	
	N	P	N	P	N	P
$\lvert I_{OFF} \rvert$ (pA)	**0.8**	**0.8**	**0.8**	**0.8**	**0.8**	**0.8**
$\lvert I_{ON} \rvert$ (µA)	**15.1**	**14.7**	**18.1**	**16.9**	**21.7**	**19.5**
$\lvert I_{EFF} \rvert$ (µA)	**7.1**	**7.1**	**8.6**	**8.2**	**10.4**	**9.5**
$\lvert V_{T,SAT} \rvert$ (V)	**0.42**	**0.42**	**0.41**	**0.40**	**0.40**	**0.39**
DIBL (mV/V)	**40**	**36**	**32**	**28**	**23**	**20**
SS_{SAT} (mV/dec)	**78**	**78**	**75**	**75**	**71**	**71**
C_{gd} (aF)	**8.1**	**7.8**	**9.2**	**8.9**	**12.0**	**11.6**
C_{gg} (aF)	**29.7**	**28.7**	**35.6**	**34.2**	**46.4**	**44.4**
$C_{gg}V_{DD}/\lvert I_{ON} \rvert$ (ps)	**1.47**	**1.45**	**1.47**	**1.52**	**1.60**	**1.70**
g_m/g_{ds}	**22.4**	**28.2**	**25.7**	**30.1**	**30.6**	**36.3**

3.4 Impact of Process-Induced Variations

3.4.1 Short Channel Effect

Fig. 3.13 shows how variations in L_g affect the saturation threshold voltage (V_T), which is defined as the gate voltage corresponding to a drain current (I_D) equal to 100 nA $\times$ W_{eff}/L_g. It can be seen that the short channel effect is reduced for iFinFETs as compared against FinFETs. Stacked-NW GAA FETs show the least amount of V_T roll-off due to their superior electrostatic integrity.

Fig. 3.13 Comparison of $V_{T,SAT}$ *vs*. L_g characteristics for multi-gate transistor technologies.

3.4.2 Impact of Inserted-Oxide Layer Thickness and Recess Amount

The sensitivities of the iFinFET performance parameters I_{ON} and intrinsic delay to variations in inserted-oxide layer thickness (T_{iox}) and recess amount are shown in **Fig. 3.14** and **Fig. 3.15**, respectively.

Fig. 3.14 Dependence of iFinFET ON-state drive current on inserted-oxide thickness (T_{iox}) and recess distance, for (a) PMOS and (b) NMOS. For reference, the ON-state drive current values for the FinFET (dashed line) and for the stacked-NW GAA FET (dotted line) are shown.

Fig. 3.15 Dependence of iFinFET intrinsic delay on inserted-oxide thickness (T_{iox}) and recess distance, for (a) PMOS and (b) NMOS. For reference, the intrinsic delay values for the FinFET (dashed line) and for the stacked-NW GAA FET (dotted line) are shown.

As was first shown in [3.6], I_{ON} increases with increasing T_{iox}. However, C_{gg} increases more quickly, resulting in increasing intrinsic delay with increasing T_{iox}. Because the recessed regions are filled with high-k dielectric material (cf. **Fig. 3.1**), which enhances I_{ON} due to greater capacitive coupling between the gate and the nanowire channel regions, a larger recess distance is beneficial for reducing intrinsic delay. It can be seen that the performance of the iFinFET is relatively insensitive to variations in

inserted-oxide layer thickness and recess amount, so that the T_{iox} requirement for the starting wafer substrate does not need to be very stringent.

3.4.3 Impact of Inserted-Oxide Layer Position

The sensitivity of the iFinFET ON-state drive current I_{ON} to variations in inserted-oxide (i-oxide) layer positions is shown in **Fig. 3.16**. For simplicity, only case of the iFinFET with one i-oxide layer is discussed herein. It can be seen that the performance of the iFinFET is relatively insensitive to variations in the positions of the single i-oxide layer, as long as the single i-oxide layer is around the center of the Si channel region above the shallow trench isolation (STI) oxide. Also, with the single i-oxide located near the center of the Si channel region above STI oxide, the best performance of the iFinFET is achieved. For the iFinFET with two i-oxide layers (cf. **Fig. 3.1**), equally distributing the two i-oxide layers in the Si channel region above STI oxide is expected to be very close to the optimal design, i.e. 1) achieving the best performance and 2) being relatively insensitive to variations in the positions of the two inserted-oxide layers.

Fig. 3.16 Dependence of NMOS iFinFET ON-state drive current on position of single inserted-oxide layer. The schematic channel cross-sections illustrate the position of the single i-oxide layer. For reference, the ON-state drive current values for the FinFET (blue dashed line), the iFinFET with two i-oxide layers (red dashed line) and the stacked-NW GAA FET (black dashed line) are also shown. W_{si} = 6 nm, H_{si} = 18 nm.

3.5 Summary

An improved evolutionary FinFET design (iFinFET) is proposed. Due to enhanced gate control achieved via fringing electric fields through inserted dielectric layers in the channel region, the iFinFET transistor design provides for improved performance and scalability as compared with the FinFET, without the need for significantly increased fabrication process complexity or a large gate capacitance penalty as for the stacked-NW GAA FET. SiO_2 is the preferred inserted-dielectric material, for low drain-induced barrier lowering. iFinFET performance is relatively insensitive to the inserted-oxide thickness, location and recess amount, and hence is robust against process-induced variations. Thus, the iFinFET is truly an intriguing new candidate transistor structure for future CMOS technologies.

3.6 References

[3.1] S. Natarajan, M. Agostinelli, S. Akbar, M. Bost, A. Bowonder, V. Chikarmane, S. Chouksey, A. Dasgupta, K. Fischer, Q. Fu, T. Ghani, M. Giles, S. Govindaraju, R. Grover, W. Han, D. Hanken, E. Haralson, M. Haran, M. Heckscher, R. Heussner, P. Jain, R. James, R. Jhaveri, I. Jin, H. Kam, E. Karl, C. Kenyon, M. Liu, Y. Luo, R. Mehandru, S. Morarka, L. Neiberg, P. Packan, A. Paliwal, C. Parker, P. Patel, R. Patel, C. Pelto, L. Pipes, P. Plekhanov, M. Prince, S. Rajamani, J. Sandford, B. Sell, S. Sivakumar, P. Smith, B. Song, K. Tone, T. Troeger, J. Wiedemer, M. Yang, and K. Zhang, "A 14 nm logic technology featuring 2nd-generation FinFET, airgapped interconnects, self-aligned double patterning and a 0.0588 μm^2 SRAM cell size," in *IEDM Tech. Dig.*, Dec. 2014, pp. 71-73.

[3.2] C. Lin, B. Greene, S. Narasimha, J. Cai, A. Bryant, C. Radens, V. Narayanan, B. Linder, H. Ho, A. Aiyar, E. Alptekin, J. An, M. Aquilino, and R. Bao, V. Basker, N. Breil, M. Brodsky, W. Chang, L. Clevenger, D. Chidambarrao, C. Christiansen, D. Conklin, C. DeWan, H. Dong, L. Economikos, B. Engel, S. Fang, D. Ferrer, A. Friedman, A. Gabor, F. Guarin, X. Guan, M. Hasanuzzaman, J. Hong, D. Hoyos, B. Jagannathan, S. Jain, S-J. Jeng, J. Johnson, B. Kannan, Y. Ke, B. Khan, B. Kim, S. Koswatta, A. Kumar, T. Kwon, U. Kwon, L. Lanzerotti, H-K Lee, W-H. Lee, A. Levesque, W. Li, Z. Li, W. Liu, S. Mahajan, K. McStay, H. Nayfeh, W. Nicoll, G. Northrop, A. Ogino, C. Pei, S. Polvino, R. Ramachandran, Z. Ren, R. Robison, I. Saraf, V. Sardesai, S. Saudari, D. Schepis, C. Sheraw, S. Siddiqui, L. Song, K. Stein, C. Tran, H. Utomo, R. Vega, G. Wang, H. Wang, W. Wang, X. Wang, D. Wehelle-Gamage, E. Woodard, Y. Xu, Y. Yang, N. Zhan, K. Zhao, C. Zhu, K. Boyd, E. Engbrecht, K. Henson, E. Kaste, S. Krishnan, E. Maciejewski, H. Shang, N. Zamdmer, R. Divakaruni, J. Rice, S. Stiffler, P. Agnello, "High Performance 14nm SOI FinFET CMOS Technology with 0.0174μm^2 embedded DRAM and 15 Levels of Cu Metallization," in *IEDM Tech. Dig*, 2014, pp. 74–76.

[3.3] S. Wu, C. Y. Lin, M. C. Chiang, J. J. Liaw, J. Y. Cheng, S. H. Yang, S. Z. Chang, M. Liang, T. Miyashita, C. H. Tsai, and C. H. Chang, V.S. Chang, Y.K. Wu, J.H. Chen, H.F. Chen, S.Y. Chang, K.H. Pan, R.F. Tsui, C.H. Yao, K.C. Ting, T.

Yamamoto, H.T. Huang, T.L. Lee, C.H. Lee, W. Chang, H.M. Lee, C.C. Chen, T. Chang, R. Chen, Y.H. Chiu, M.H. Tsai, S. M. Jang, K.S. Chen, Y. Ku, "An Enhanced 16nm CMOS Technology Featuring 2nd Generation FinFET Transistors and Advanced Cu/ low-k Interconnect for Low Power and High Performance Applications," in *IEDM Tech. Dig*, 2014, pp. 48–51.

[3.4] "International Technology Roadmap for Semiconductors (ITRS), 2013." [Online]. Available: http://public.itrs.net/.

[3.5] C. Dupré, A. Hubert, S. Becu, M. Jublot, V. Maffini-Alvaro, C. Vizioz, F. Aussenac, C. Arvet, S. Barnola, J.-M. Hartmann, G. Garnier, F. Allain, J.-P. Colonna, M. Rivoire, L. Baud, S. Pauliac, V. Loup, T. Chevolleau, P. Rivallin, B. Guillaumot, G. Ghibaudo, O. Faynot, T. Ernst, and S. Deleonibus, "15nm-diameter 3D Stacked Nanowires with Independent Gates Operation: ΦFET," in *IEDM Tech. Dig*, 2008, pp. 1 – 4.

[3.6] P. Zheng, D. Connelly, F. Ding and T.-J. King Liu, "Inserted-oxide FinFET (iFinFET) Design to Extend CMOS Scaling," in *VLSI-TSA Symp*., 2015, pp. 1-2.

[3.7] P. Zheng, D. Connelly, F. Ding and T.-J. King Liu, "Simulation-Based Study of the Inserted-Oxide FinFET for Future Low-Power System-on-Chip Applications," *IEEE Electron Device Letters*, vol. 36, no. 8, pp. 742-744, 2015.

[3.8] P. Zheng, D. Connelly, F. Ding, T.-J. King Liu "FinFET Evolution Toward Stacked-Nanowire FET for CMOS Technology Scaling", *IEEE Transactions on Electron Devices,* vol 62, no. 12, pp. 3945 – 3950, 2015.

[3.9] Sentaurus User's Manual, Synopsys, Inc., Version J-2014.09-SP1.

[3.10] A. Brand, "Precision Materials to Meet Scaling Challenges Beyond 14nm", SEMICON West, 2013.

[3.11] R. Divakaruni, "SOI CMOS Technology Through 7nm", Silicon on Insulator Technology Summit, 2013.

[3.12] K.-H. Kao, A. S. Verhulst, W. G. Vandenberghe, B. Soree, G. Groeseneken, and K. de Meyer, "Direct and indirect band-to-band tunneling in germanium-based TFETs," IEEE Trans. Electron Devices, vol. 59, no. 2, pp. 292–301, 2012.

[3.13] M. G. Ancona, H. F. Tiersten, "Macroscopic physics of the silicon inversion layer," Physical Review B, vol. 35, 15, pp.7959-7965, 1987.

[3.14] M. Ancona, "Density-gradient theory: a macroscopic approach to quantum confinement and tunneling in semiconductor devices," Journal of Computational Electronics, vol.10 pp. 65-97, 2011.

[3.15] L. Chang, M. Ieong and M. Yang, "CMOS Circuit Performance Enhancement by Surface Orientation Optimization," IEEE Trans. Electron Devices, vol. 51, no. 10, pp. 1621–1627, 2004.

Chapter 4

Sub-lithographic Patterning via Tilted Ion Implantation (TII)

4.1 Introduction

Advancements in lithography have been key to sustaining Moore's Law. Due to the low transmittance of blank mask materials and/or the availability of high-intensity light sources for wavelengths shorter than 193 nm [4.1-4.2], the semiconductor industry has resorted to "multiple-patterning" techniques to increase the density of linear features patterned on a chip, since the 22 nm CMOS technology node [4.3-4.4]. The most two commonly used multiple-patterning techniques in high-volume manufacturing (HVM) "SADP" and "LELE" have been briefly reviewed in Chapter 1. SADP (cf. **Fig. 1.6**) involves multiple thin-film deposition steps as well as multiple dry-etch steps. Thus, it has significant process complexity and associated cost. LELE (cf. **Fig. 1.7**), implied by its name, involves roughly twice the number of processes as the conventional process to pattern a single integrated circuit (IC) layer. Overlay and line-width-roughness (LWR) issues have also prevented LELE from being used for critical layers.

Several "next-generation lithography" techniques have also been extensively investigated in recent years, but quite a few significant technical challenges hinder their practical application in HVM. Extreme ultraviolet (EUV) lithography is confirmed to be not ready sooner than the early stages of the 7 nm technology node, due to challenges of achieving a high-power light source, mask defect detection, etc. [4.5-4.6]. Directed self-assembly (DSA) technique [4.7] have drawbacks including a limited range of feature sizes and pitches for a given diblock copolymer material formulation, and feature-edge roughness which does not scale well with the feature size. Nanoimprint lithography (NIL) [4.8] requires a high-resolution template and stringent defect control.

Hence, there remains an urgent need [4.9] to develop a cost-effective sub-lithographic patterning technique for scaling down the minimum feature size and increasing the density of features in an IC "chip," to sustain Moore's Law.

Tilted ion implantation (TII) can be used to from sub-lithographic features [4.10-4.12]. This chapter discusses TII-enhanced patterning as a promising cost-effective patterning technique to facilitate CMOS technology scaling to sub-7 nm nodes. The general concept is the use of ion implantation to enhance the etch rate of a thin masking layer and to perform the implantation at tilted angles to achieve sub-lithographic implanted regions that are self-aligned to pre-existing photoresist or hard-mask features over the masking layer on the surface of the IC layer to be patterned. A TII-based pitch-halving process flow is described. The patterning resolution limit of TII is investigated via experiments as well as rigorous Monte Carlo process simulations. TII-enhanced

patterning is feasible for IC layers in both front-end-of-line (FEOL) and low-temperature back-end-of-line (BEOL) processes. With feature size below 10 nm experimentally demonstrated, TII-enhanced patterning offers a cost-effective pathway to extend the era of Moore's Law.

4.2 TII-Enhanced Patterning Concept

Fig 4.1 illustrates the TII approach for doubling the density of linear features in an IC layer. After a thin oxide layer is formed over the IC layer, a photo-resist (PR) or hard-mask (HM) layer is deposited and patterned. Then ion implantation is performed at positive tilt angle and also at negative tilt angle to selectively damage the oxide layer in regions self-aligned to the PR or HM features, leaving the central region between the PR or HM features unimplanted due to the shadowing effect. Since the wet-etch rate of implanted SiO_2 can be significantly enhanced due to structural damage [4.13], the implanted regions of the oxide layer can be selectively removed in dilute hydrofluoric (DHF) acid. Afterwards the patterned oxide layer can be used to mask the etching of the underlying IC layer, resulting in twice as many features (trenches) as in the PR or HM layer.

PR or HM
PR or HM
PR or HM
PR or HM
(a)
OXIDE
IC LAYER
(b)
(c)
(d)
(e)
(f)

Fig. 4.1 Schematic cross-sections illustrating the tilted ion implantation approach for pitch-halving. Key steps: (b) ion implantation is performed at positive (black dashed lines) tilt angle and also at negative (red dashed lines) tilt angle to selectively damage regions of the oxide layer self-aligned to the pre-existing photoresist (PR) or hard-mask (HM) features, leaving the central region between the PR or HM features undamaged due to the shadowing effect; (c) portions of the oxide layer which are damaged are etched away more rapidly than undamaged portions. (e) the patterned oxide layer is then used as a hard mask itself, to pattern the underlying IC layer.

An initial experiment was conducted to prove the concept of this enhanced patterning approach, as follows: Photoresist lines were formed using conventional lithography (a deep-ultraviolet (DUV) stepper with 248 nm excimer laser) on the surface

of a Si wafer coated with a thin (20 nm-thick) SiO_2 (450 °C low-temperature-oxide) hard mask layer. Argon ion (Ar^+) implantation (dose = 7.0×10^{13} cm^{-2}, energy = 13.8 keV) was performed only at a single tilt angle of 41° to selectively damage the SiO_2 in regions not (shadow)masked by the patterned photoresist. Afterwards, a sample of the wafer was dipped in dilute buffered-HF solution to selectively remove the implanted regions of the SiO_2 hard mask layer without attacking the photoresist. Subsequently the sample was subjected to a Si etch process. Due to the relatively poor selectivity of this particular process (<10:1 Si:SiO_2 etch-rate ratio), the exposed SiO_2 was eventually etched away during this process so that the Si was etched everywhere between the photoresist lines. Nevertheless, sub-lithographic features (trenches in the Si) corresponding to the implanted regions between the photoresist lines can be distinguished in the cross-sectional scanning electron micrograph (XSEM) of **Fig. 4.2**.

Fig. 4.2 Cross-sectional SEM image of photoresist lines on the surface of a Si wafer which was etched to form sub-lithographic features corresponding to the regions in which the thin SiO_2 hard mask layer was damaged by only negative-angle tilted ion implantation.

To demonstrate TII-enhanced patterning with further scaled dimensions, amorphous silicon (a-Si) is chosen to replace photoresist for the pre-patterned HM layer before ion implantation. Spacer lithography was used in conjunction with conventional 248 nm DUV lithography to form a-Si hard-mask features with sub-100 nm dimensions over a thermally oxidized (10 nm-thick SiO_2) silicon wafer substrate. A single tilted Ar^+ implant (dose = 3×10^{14} cm^{-2}, energy = 3 keV) was used to define sub-lithographic trenches in the surface of the Si wafer substrate, using this 10 nm-thick SiO_2 for the thin oxide masking layer. After the implanted regions of the SiO_2 layer were selectively removed by DHF treatment, a reactive ion etching (RIE) process was used to selectively etch the exposed regions of crystalline Si (c-Si), making it easy to distinguish TII-defined regions. The XSEM in **Fig. 4.3** clearly shows that the unimplanted SiO_2 (unetched c-Si) regions are very uniform in width. The plan-view SEM images in **Fig. 4.4** show that the feature edges defined by TII are self-aligned to the a-Si HM edges, reproducing with high fidelity the line-edge roughness (LER) of the a-Si HM features.

Fig. 4.3 Cross-sectional SEM image showing sub-lithographic trenches patterned by only positive-angle tilted ion implantation. ①: a-Si hard mask, ②: c-Si (unetched due to SiO_2 protection), ③: sub-lithographic trench in c-Si, ④: thin SiO_2 layer underneath a-Si masking feature.

Fig. 4.4 Plan-view SEM images showing that the features defined by tilted ion implantation are self-aligned to the lithographically patterned a-Si hard mask features, reproducing with high fidelity the line-edge roughness of the hard mask, for (a) 15° implantation tilt angle, and (b) 20° implantation tilt angle. ①: a-Si hard mask, ②: c-Si (unetched, due to SiO_2 protection), ③ Sub-lithographic trench in c-Si.

4.3 Demonstration of Pitch-halving via TII Patterning

To demonstrate the feasibility of pitch-halving, doubly tilted Ar^+ implantation (*i.e.*, tilt angle = ±15°) with 3.0 keV acceleration energy and 3×10^{14} cm^{-2} dose was performed on another sample. Subsequently, the sample was subjected to the dilute buffered-HF wet etching and Si dry etch processes described in Chapter 4.2. **Fig. 4.5** shows that the TII patterning is capable of doubling the density of features, as proposed. Considering that the distance between a-Si hard-mask features is ~64 nm, the achieved effective (local) pitch of the etched c-Si features is ~42 nm (2/3 × 64 nm). Therefore, ~21 nm (local) local half-pitch feature is achieved by TII double patterning.

Fig. 4.5 Cross-sectional SEM images demonstrating the feasibility of doubly tilted ion implantation as a double-patterning approach (cf. Fig. 4.1).

4.4 TII Patterning Resolution Limit

4.4.1 Monte Carlo Process Simulations

Herein the resolution limit of the TII patterning technique is systematically investigated via Monte Carlo process simulations using Sentaurus [4.14] to obtain the three-dimensional (3-D) damage profile resulting from tilted Ar^+ implantation. The ion acceleration energy and dose were adjusted to achieve the same projected range and peak damage concentration for each tilt angle, based on SRIM simulations (**Fig. 4.6**) [4.15].

Fig. 4.6 Simulated Ar^+ implantation-induced damage depth profile obtained using SRIM [4.15]. For different values of the implant tilt angle (θ), the implant energy (E) and dose (D) are adjusted to achieve the same projected range and peak damage value, respectively, in the SiO_2 layer.

Fig. 4.7 shows the simulated 3-D structure, comprising perfectly linear a-Si HM features formed on top of an oxidized Si substrate. The thickness of the SiO_2 layer is 5 nm. It should be noted that, in order to pattern fine features in the SiO_2 masking layer, this layer thickness should be very thin to avoid significant lateral etching during the post-implantation wet etch process.

Fig. 4.7 (a) Isometric view, (b) plan view, and (c) cross-sectional view of the 3-D structure used for Monte-Carlo ion implantation simulations, using Sentaurus [4.14]. The hard mask (HM) is a patterned layer of a-Si.

Fig. 4.8 shows the impact of variations in HM corner radius on the average size and latent LER of the implanted oxide region. It can be seen that a very small feature (~16 nm-wide implanted oxide region, in this case) can be defined by TII and that the latent LER introduced by TII is relatively small and insensitive to variations in HM corner radius.

Fig. 4.8 Simulated impact of hard mask (HM) corner radius on the average feature size (F) and latent line-edge roughness (LER) of the implanted oxide region. Error bars represent +/- 1 standard deviation. Implant tilt angle θ= 15°.

Different feature sizes can be achieved simply by using different implant tilt angles, which offers flexibility in design. **Fig. 4.9** shows the impact of implant tilt angle (θ) on the latent LER of the implanted oxide region. Due to increased lateral implant straggle, LER increases with the tilt angle. For $\theta < 30°$, lateral straggle is not anticipated to limit the resolution of the TII patterning technique.

Fig. 4.9 Simulated impact of implant tilt angle θ on the latent line-edge roughness (LER) of the implanted oxide region. Error bars represent +/- 1 standard deviation.

Fig. 4.10 shows cross-sectional views of the simulated structure for two different implant tilt angles. The higher concentration of ions implanted into the HM sidewall for higher tilt angle (θ = 30°) indicates that ions are more prone to be backscattered by HM sidewall for lower tilt angle (θ = 15°).

Fig. 4.10 Cross-sectional view of the 3-D structure used for Monte-Carlo ion implantation simulations, using Sentaurus [4.14]. The hard mask (HM) is a patterned layer of a-Si. The Ar concentration in the HM is represented in color using a linear scale, indicating that Ar ions are more prone to be backscattered from the HM sidewall for shallower tilt angle ($\theta = 15°$). For $\theta = 15°$, E = 1.50 keV, D = 2.00×10^{14} cm^{-2}; for $\theta = 30°$, E = 1.87 keV, D = 1.69×10^{14} cm^{-2} (cf. Fig. 4.6), ensuring the same projected range and peak damage concentration in the SiO_2 layer, for each tilt angle. HM pitch = 125 nm.

To decouple ions received/backscattered by the HM top surface and sidewall, a set of simulations for different HM heights was performed (**Fig. 4.11**). Assuming that a change in the HM height (> 30 nm) does not change the amount of ions implanted into the HM top surface, linear extrapolation of the data in **Fig. 4.11** will give the number of ions implanted into the HM top surface. Subtracting this value from the total number of ions inside the HM results in the number of ions that are implanted into the HM sidewall.

The tilt-angle design tradeoff between low back-scattering rate and low latent LER (determined by lateral implant straggle) for large HM pitch is shown in **Fig 4.12**. Back-scattering can result in increased latent LER for very small HM pitch, as shown in **Fig. 4.13**. Based on these simulation results, TII is projected to be suitable for patterning features as small as 10 nm with good fidelity.

Fig. 4.11 Simulated total number of ions implanted into the HM, as a function of HM height. By extrapolation to zero HM height, the number of ions implanted into the HM top surface can be determined, and consequently the number of ions implanted into the HM sidewall.

Fig. 4.12 Simulated impact of implant tilt angle (θ) on the ion backscattering rate (left y-axis) and the latent line-edge roughness (LER) of the implanted oxide region (right y-axis). For θ = 15°, E = 1.50 keV, D = 2.00×10^{14} cm^{-2}; for θ = 30°, E = 1.87 keV, D = 1.69×10^{14} cm^{-2} (cf. Fig. 4.6), ensuring the same projected range and peak damage concentration in the SiO_2 layer, for each tilt angle. HM pitch = 125 nm.

Fig. 4.13 Simulated impact of ion backscattering from the HM sidewall, on the latent line-edge roughness (LER) of the implanted oxide region. (The average implant-defined feature size, F, is changed by adjusting the HM pitch, and the HM corner radius is fixed at 3 nm.) Tilted ion implantation is projected to be suitable for patterning features as small as 10 nm with good fidelity. Error bars represent +/- 1 standard deviation.

4.4.2 Experimental Investigation of Line Edge Roughness

Fig. 4.14 shows a higher-magnification plan view of HM and TII-defined feature edges. It should be noted that the roughness of the TII-defined feature edge is affected not only by the latent LER of the implanted oxide region, but also by post-implant wet etching of the SiO_2 masking layer and subsequent dry etching of the c-Si substrate. As a result, the standard deviation from the mean position of the TII-defined edge is smaller than that of the HM edge, as indicated in **Fig. 4.14 (a)**. (The LER analysis was performed using the SuMMIT software package [4.16].) **Fig. 4.14 (b)** compares the edge deviations for the TII-defined edge and the HM edge along the length of the feature, *i.e.* in the vertical direction in **Fig. 4.14 (a)**. The self-aligned nature of TII patterning is evident from the fact that the TII-defined edge closely tracks the HM edges with a correlation coefficient of 0.91 as shown in **Fig. 4.14 (c)**.

Fig. 4.14 (a) Plan-view SEM image showing that the line-edge roughness (LER) of a TII-defined edge is lower than that of its corresponding HM edge. ①: a-Si hard mask, ②: c-Si (unetched, due to SiO_2 protection), ③ Sub-lithographic trench in c-Si. (b) deviations (from average position) of TII-defined edge and HM edge along the length of the masking feature (*i.e.* in the vertical direction in Fig. 4.14(a)). (c) correlation between deviations of TII-defined edge and HM edge.

Due to the stochastic nature of the ion implantation process, higher spatial frequency components of LER can be more significant for TII-defined edges. This is verified in **Fig. 4.15** which compares the power spectral density (PSD) of TII-defined edges *vs.* HM edges. Note that TII improves low- and mid-frequency LER, which has been identified as a major challenge for the industry [4.17-4.18]. The plan-view SEM in **Fig. 4.16** demonstrates that features down to 9 nm can be defined by the TII technique.

Fig. 4.15 Power spectral density comparison for TII-defined edges *vs.* HM edges (36 samples, each 2.74 μm in length).

Fig. 4.16 Plan-view SEM image demonstrating that feature size down to **9 nm** can be achieved via tilted ion implantation patterning in a self-aligned manner, reproducing with high fidelity the line-edge roughness of the hard mask. ①: a-Si hard mask, ②: c-Si (unetched, due to SiO_2 protection), ③ Sub-lithographic trench in c-Si.

4.5 Summary

Tilted ion implantation (TII) patterning is an effective technique for forming sub-lithographic features and for increasing the density of features on a chip. It can be used to pattern features with dimensions below 10 nm and with lower line-edge roughness than that of pre-existing masking features on the surface of a substrate, in a self-aligned manner. Since ion implantation is a well-established technique with good process control and high throughput, the TII double-patterning approach can be easily adopted in high-volume manufacturing processes and shows promise for extending IC technology advancement beyond the 7 nm technology node (sub-40 nm pitch).

4.6 References

[4.1] C. Wagner, N. Harned, "EUV Lithography: Lithography Gets Extreme," *Nature Photonics* 2010, 4, pp. 24-26.

[4.2] C. Willson, B. Roman, "The Future of Lithography: SEMATECH Litho Forum 2008," *ACS Nano* 2008, pp. 1323-1328.

[4.3] "International Technology Roadmap for Semiconductors (ITRS), 2013." [Online]. Available: http://public.itrs.net/.

[4.4] C. Auth, C. Allen, A Blattner, D. Bergstrom, M. Brazier, M. Bost, M. Buehler, V. Chikarmane, T. Ghani, T. Glassman, R. Grover, W. Han, D. Hanken, M. Hattendorf, P. Hentges, R. Heussner, J. Hicks, D. Ingerly, P. Jain, S. Jaloviar, R. James, D. Jones, J. Jopling, S. Joshi, C. Kenyon, H. Liu, R. McFadden, B. McIntyre, J. Neirynck, C. Parker, L. Pipes, I. Post, S. Pradhan, M. Prince, S. Ramey, T. Reynolds, J. Roesler, J. Sandford, J. Seiple, P. Smith, C. Thomas, D. Towner, T. Troeger, C. Weber, P. Yashar, K. Zawadzki, and K. Mistry, "A 22nm high performance and low-power CMOS technology featuring fully-depleted tri-gate transistors, self-aligned contacts and high density MIM capacitors," *VLSI Technol. (VLSIT), 2012 Symp*., pp. 131–132, 2012.

[4.5] B. Turkot, S. Carson, A. Lio, T. Liang, M. Phillips, B. McCool, E. Stenehjem, T. Crimmins, G. Zhang, G. Sivakumar, "EUV Progress Toward HVM Readiness", *Proc. SPIE 9776, Extreme Ultraviolet (EUV) Lithography VII*, 2016, 9776, 977602-01-977602-09.

[4.6] A. Pirati, R. Peeters, D. Smith, S. Lok, M. Noordenburg, R. Es, E. Verhoeven, H. Meijer, A. Minnaert, J.-W. Horst *et al.*, "EUV Lithography Performance for Manufacturing: Status and Outlook," Proc. *SPIE 9776, Extreme Ultraviolet (EUV) Lithography VII*, 2016, 97760A-1- 97760A-15.

[4.7] C. T. Black, R. Ruiz, G. Breyta, J. Y. Cheng, M. E. Colburn, K. W. Guarini, *et al.*, "Polymer self assembly in semiconductor microelectronics," *IBM J. Res. Dev.,* vol. 51, pp. 605-633, 2007.

[4.8] Malloy, M. and Litt, L. C., "Technology review and assessment of nanoimprint lithography for semiconductor and patterned media manufacturing," *J. Micro/Nanolith. MEMS MOEMS*, 10(3), 032001, 2011.

[4.9] M. Lapedus, "Will 5nm Happen?" *Semi. Engineering*, 2016.

[4.10] R. Divakaruni, S. Kudelka, H. Tews, I. McStay, K.-H. Lee, U. Schroeder, U.S. Patent No. 6,498,061 (2000).
[4.11] P.-I. Lee, ; C.-Y. Lee, C.-L. Cheng, U.S. Patent No. 7,316,978 (2005).
[4.12] S. W. Kim, P. Zheng, K. Kato, L. Rubin, T.-J. King Liu, "Enhanced Patterning by Tilted Ion Implantation'" *Proc. SPIE Alternative Lithographic Technologies VIII,* 2016, 9777, 97771B1-97771B6.
[4.13] X. Sun, Q. Lu, H. Takeuchi, S. Balasubramanian, T.-J. King Liu, "Selective Enhancement of SiO_2 Etch Rate by Ar Ion Implantation for Improved Etch Depth Control," *Electrochemical and Solid-State Letters* 2007, 10(9), pp. 89-91.
[4.14] Sentaurus User's Manual (Synopsys, Inc.), Ver. J-2015.06.
[4.15] http://www/srim.org.
[4.16] www.lithometrix.com
[4.17] E. Gogolides, V. Constantoudis, G. Kokkoris, "Toward an Integrated Line Edge Roughness Understanding: Metrology, Characterization, and Plasma Etching Transfer", *Proc. SPIE 8685, Advanced Etch Technology for Nanopatterning II* 2013, 8685, 868505-01-868505-16
[4.18] N. Mohanty, R. Farrell, C. Periera, K. Subhadeep, E. Franke, E. Smith, A. Ko, A. DeVilliers, P. Biolsi, L. Sun *et al.*, LER Improvement for Sub-32nm Pitch Self-Aligned Quadruple Patterning (SAQP) at Back End Of Line (BEOL). *Proc. SPIE 9782, Advanced Etch Technology for Nanopatterning V* 2016, 9782, 97820Q-1-97820Q-15.

Chapter 5

IC Manufacturing Cost Analyses

5.1 Introduction

As indicated in Moore's 1965 paper [5.1], "reduced cost is one of the big attractions of integrated electronics, and the cost advantage continues to increase as the technology evolves toward the production of larger and larger circuit functions on a single semiconductor substrate" (cf. **Fig. 1.2**). Indeed, the primary reason for increasing the number of components per IC, enabled by advancement of IC manufacturing technology, was (and still) is lower cost. However, significant and escalating cost challenges are seen by the semiconductor industry in advanced technology nodes (**Fig. 5.1**), which is mostly due to the adoption of 3-D FinFET structures and multiple-patterning techniques. The issue of increasing cost is identified as the focus of concerns and doubts over the vitality of Moore's Law going forward [5.2].

Fig. 5.1 Foundry Wafer Cost Trend. Source: IC Knowledges Cost Model [5.3]. Note: assumptions for 7 nm node are made that EUV is available and EUV has the potential to decrease the wafer cost versus 10 nm node.

This chapter firstly summarizes a survey regarding IC manufacturing cost throughout the semiconductor industry. Publicly available data and perspectives are collected from representatives of integrated device manufacturers (IDM), foundries, fabless companies, and industry observers. Then two case studies of wafer processing

cost are presented: one for FinFET *vs.* iFinFET *vs.* stacked-NW GAA MOSFET, and the other for SADP *vs.* TII double patterning. The results show that the iFinFET technology and TII double patterning technique have significant economic merit.

5.2 Manufacturing Cost Challenges for Advanced CMOS Technologies

It is well known that different companies are concerned with different IC manufacturing process complexities and use different transistor densities in their products, even at the same technology node. Thus, to make a fair comparison, cost-per-transistor is chosen as the figure-of-merit to benchmark cost, as was done in Moore's 1965 paper [5.1]. It should be noted this work only discusses semiconductor manufacturing cost. Consideration of IC chip price (relevant for consumers) is beyond the scope of this work.

Fig. 5.2 shows Intel Corporation's perspective on cost-per-transistor. Although the cost-per-area of an IC chip has been increasing as technology advances over time, by aggressively increasing transistor density the cost-per-transistor can be lowered, noticeably at a higher-than historical-pace since the 14 nm technology node.

Fig. 5.2 Intel's "offsetting wafer cost with density" strategy [5.2]. Source: Intel Corp.

However, it is observed and projected by International Business Strategies (IBS) that cost-per-transistor reaches a minimum at the 28 nm technology node and increases for the 20 nm and 16/14 nm technology nodes (**Fig. 5.3**). This perspective is echoed quite often in the industry, but it should be noted that majority of the companies in the industry are fabless. Moreover, IBS later clarifies that this analysis accounts for logic products, but not memory products [5.4].

Fig. 5.3 IBS' cost per gate trend [5.5]. Source: International Business Strategies, Inc.

Believing that the real scenario falls in-between the above-mentioned two cases, ARM Research illustrates that cost-per-transistor reduces below 28 nm node, but at a slower rate (**Fig. 5.4**). It should also be noted that ARM's analysis explicitly includes cost due to non-recurring engineering (NRE) costs, which comprise mask set cost and design cost [5.6].

Fig. 5.4 ARM's cost-per-transistor scaling trend, including NRE [5.6]. Source: ARM Research.

In **Fig. 5.5**, IC Knowledge illustrates the foundry cost-per-transistor trend [5.3]. Unlike IBS, it shows that the cost-per-transistor at the 20 nm node is lower than that at the 28 nm node. The increase in cost-per-transistor at the 16 nm node is attributed to the lack of a “shrink” in the foundry fabrication process. At 7 nm, the assumption of a significant reduction in lithographic mask count is made due to the advent of extreme ultra-violet (EUV) lithography, which is potentially capable of reducing overall cost-per-wafer (cf. **Fig. 5.1**)

Fig. 5.5 Foundry cost per gate trend [5.3]. Source: IC knowledges Strategic Cost Model.

Although different statements have been made regarding the cost-per-transistor trend (due to different assumptions), there is consensus that the industry truly needs cost-effective technologies for scaling beyond the 7 nm technology node.

5.3 Manufacturing Cost Comparison of FinFET *vs.* iFinFET *vs.* GAA MOSFET

It was mentioned in **Chapter 3.1** that the iFinFET can be fabricated using a process identical to that for a conventional bulk FinFET but starting with a multi-SOI substrate (silicon-on-insulator on silicon-on-insulator). The stacked-NW GAA MOSFET, on the other hand, requires significantly different processes. It requires epitaxial growth

of relatively thick (>10 nm) $Si_{1-x}Ge_x$ sacrificial layers between Si channel layers to accommodate gate-dielectric/gate-metal/gate-dielectric layers in-between the NWs, so that fin structures with very high aspect ratio must be etched prior to selective removal of the sacrificial $Si_{1-x}Ge_x$ layers. Some key processes of fabricating a stacked-NW GAA MOSFET are illustrated in **Fig. 5.6**.

Fig. 5.6 Cross-sectional views across the channel region of stacked-NW GAA MOSFET illustrate the key processing steps of its fabrication. (Left): Relatively thick (>10 nm) $Si_{1-x}Ge_x$ sacrificial layers between Si channel layers are epitaxially grown. The whole structure is then etched into a fin structure with very high aspect ratio. (Middle): $Si_{1-x}Ge_x$ sacrificial layers are selectively removed in the channel region to make room for gate-dielectric/gate-metal/gate-dielectric layers. (Right): Gate-dielectrics and gate-metal are formed by atomic layer deposition (ALD), respectively.

Table 5.1 compares the additional (to FinFET) processes and cost associated [5.7-5.8] with the iFinFET and the stacked-NW GAA MOSFET. At first glance, it may seem that fabrication of the stacked-NW GAA MOSFET is more cost-efficient than that of the iFinFET. However, the calculation is based on the assumption that the two fabrication flows have the same yield. In reality, iFinFET can leverage the already established FinFET fabrication flow, but it is overly optimistic to assume that the fabrication of the stacked-NW GAA MOSFET can achieve comparable yield due to its complex device structure. It should be noted that high-quality (low-defect-density) ultra-thin-buried-oxide SOI wafers are available for high-volume CMOS manufacturing today; there is no technical reason why high-quality multi-SOI wafers cannot be made in the future if there is demand for them.

Table 5.1. ADDITIONAL PROCESSES AND COST ASSOCIATED WITH "iFinFET" AND "stacked-NW GAA MOSFET" TO "FinFET"

iFinFET				Stacked-NW GAA MOSFET			
Process Steps		Cost (/wafer)		Process Steps		Cost (/wafer)	
Process	Description	Min.	Max.	Process	Description	Min.	Max.
Wafer bonding	1st SOI	\$40	\$58	Epitaxy	$Si_{1-x}Ge_x$	\$20	\$30
Wafer bonding	2nd SOI	\$40	\$58	Epitaxy	$Si/Si_{1-x}Ge_x/Si$	\$28	\$42
Dry etch	2 i-oxide layers	\$8	\$11	Dry etch	2 $Si_{1-x}Ge_x$ layers	\$8	\$11
				Wet etch	$Si_{1-x}Ge_x$ removal	\$2	\$4
	Total:	\$88	\$127		Total:	\$58	\$87

It should also be noted that the iFinFET can be fabricated in a similar fashion that stacked-NW GAA MOSFET is fabricated. In this case, relatively thin sacrificial $Si_{1-x}Ge_x$ layers are needed, as they only need to make room for the inserted-oxide layers but not the metal layers as in the stacked-NW GAA MOSFET. In addition, gate stack formation will be relatively easier as well, since iFinFET essentially is a "tri-gate" device. Nevertheless, this approach is still expected to have yield issue, which can be avoided by using the proposed multi-SOI substrate.

Indeed, adoption of the iFinFET or the stacked-NW GAA MOSFET makes the total wafer cost increase. However, transistor gate length (and thus gate pitch) can be shorter in iFinFET or stacked-NW GAA MOSFET technology, thanks to the enhanced gate control. Hence, transistor density is higher in iFinFET or stacked-NW GAA MOSFET technology. Considering logic wafer cost (> \$4000) in 16 nm FinFET process [5.5], [5.8] and even higher cost in future technology nodes, the additional wafer cost (< 2.5%) brought by iFinFET or stacked-NW GAA MOSFET technology is expected to be offset by the transistor density benefit. Thus, cost-per-transistor can still be lower for iFinFET or stacked-NW GAA MOSFET technology than for FinFET technology.

5.4 Manufacturing Cost Comparison of SADP *vs.* TII Double Patterning

Lithography usually contributes much to the economic burden associated with technology advancement. This is especially true in advanced technology nodes where multiple patterning is the key reason for cost-reduction challenge. Indeed, the sheer cost and complexity of multiple patterning could dissuade chipmakers from jumping to future nodes, thereby stunning the growth rates of the IC industry [5.9]. Considering that the long-waited EUV has not yet proved its readiness for high-volume manufacturing (HVM), SADP is the only HVM ready patterning technique for critical layers of IC processes. **Chapter 4** demonstrates that TII-enhanced patterning is a technologically viable solution to extend IC technology advancement beyond the 7 nm technology node. **Table 5.2** and **Fig. 5.7** compare the number of process steps and costs [5.7], [5.10] associated with TII double patterning, against those of SADP. The result shows that TII double patterning involves fewer steps and does not require new or aggressive process capabilities. Thus, TII double patterning is much more cost-efficient than SADP.

Table 5.2. COST COMPARISON OF DOUBLE-PATTERNING APPROACHES

Self-aligned Double Patterning (SADP)				Tilted Ion Implantation (TII) Double Patterning			
Process Steps		Cost (/wafer)		Process Steps		Cost (/wafer)	
Process	Description	Min.	Max.	Process	Description	Min.	Max.
Oxide formation	Etch stopper	$1.5	$3	Oxide formation	Masking layer	$1.5	$3
CVD	Mandrel layer	$4	$5	Photolithography	Lithography	$20	$60
Photolithography	Lithography	$20	$60	Ion implantation	1st tilt	$0.53	$2
Dry etch	Mandrel etch	$8	$11	Ion implantation	2nd tilt	$0.32	$1.2
ALD	Sidewall	$15	$30	Wet etch	Selective etch	$2	$4
Dry etch	spacers	$8	$11	Dry etch	Pattern transfer	$8	$11
Dry etch	Mandrel pull	$8	$11	Wet etch	SiO_2 removal	$2	$4
Dry etch	Pattern transfer	$8	$11				
Dry etch	Pattern transfer	$8	$11				
Wet etch	Spacer removal	$2	$4				
Wet etch	Etch stopper removal	$2	$4				
	Total:	$84.5	$161		Total:	$34.4	$85.2

Fig. 5.7 Comparison of minimum (min.) process costs for double patterning approaches: (left) SADP and (right) TII double patterning. Process step abbreviations (used for the x-axes): O (oxide formation), C (CVD – chemical vapor deposition), P (photolithography), D (dry etch), A (ALD – atomic layer deposition), W (wet etch) and I (ion implantation).

5.5 Summary

Although different opinions are held throughout industry regarding the “cost-per-transistor” trend, reduction in IC manufacturing cost is the key challenge as technology advances to extend Moore’s Law. The iFinFET technology and TII double patterning

technique show promise for cost reduction in future technology nodes, especially beyond the 7 nm technology node where the industry does not yet have clear solutions.

5.6 References

[5.1] G. E. Moore, "Cramming more components onto integrated circuits", Electronics, volume 38, number 8, April 19, 1965, pp.114-117.

[5.2] W. M. Holt, "Moore's Law: A Path Forward", *2016 International Solid State Circuits Conference*, paper 1.1.

[5.3] S. Jones, "Moore's Law is dead, long live Moore's Law", *SemiWiki*, April 21th, 2015. [Online].

[5.4] H. Jones, "10 nm Chips Promise Lower Costs", *EETimes*, June 15th 2015. [Online].

[5.5] H. Jones, "Why migration to 20nm bulk CMOS and 16/14nm finFETs is not best approach for semiconductor industry," January 2014. [Online]. Available: http://www.ibs-inc.net/.

[5.6] G. Yeric, "Moore's Law at 50: Are we planning for retirement?" in *IEDM Tech. Dig.*, Dec. 2015, pp. 1.1.1-1.1.8.

[5.7] B.-Y. Nguyen, private communication.

[5.8] Z. Or-Bach, B. Cronquist, I. Beinglass, Z. Wurman, "How much does ion-cut cost?" July 25th 2011. [Online]. Available: http://www.monolithic3d.com/

[5.9] M. Lapedus, "What if EUV fails?" *Semi. Engineering*, April 17th, 2014.

[5.10] L. Rubin, private communication.

Chapter 6
Conclusion

6.1 Contributions of This Work

The semiconductor industry has made great innovations to sustain the cadence of Moore's Law [6.1]. To control short channel effects (SCE) and to suppress high OFF-state leakage current, the 3-D FinFET transistor structure has been adopted since the 22 nm technology node [6.2]. To keep increasing the density of linear features patterned on an IC chip, "multiple-patterning" techniques also have been used in high-volume manufacturing (HVM) since the 22 nm technology node [6.2]. It's generally believed that the pace of IC technology advancement will slow down dramatically in the sub-20 nm (minimum half-pitch) regime. Considering that the industry does not yet have clear solutions for beyond the 7 nm technology node [6.3], innovations in transistor design and fabrication processes are needed to address this issue.

This work firstly compares the performance and variability of a GAA MOSFET against that of an ideal silicon-on-insulator (SOI) FinFET design at 10 nm gate length (anticipated for 4/3 nm CMOS technology), accounting for systematic and random variations [6.4]. The benefits of GAA MOSFET technology for lowering the minimum operating voltage (V_{min}) and area of a six-transistor (6-T) SRAM cell to facilitate increased transistor density following Moore's Law are assessed.

To circumvent the significant added fabrication process complexity and a large gate capacitance penalty as for a stacked-NW GAA MOSFET, an improved evolutionary FinFET design, the inserted-oxide FinFET (iFinFET), is discussed in this work for scaling beyond the FinFET [6.5-6.6]. Due to enhanced gate control achieved via fringing electric fields through inserted dielectric layers in the channel region, the iFinFET transistor design provides for improved performance and scalability as compared with the FinFET. SiO_2 is found to be the preferred inserted-dielectric material, for low drain-induced barrier lowering. iFinFET performance is relatively insensitive to the inserted-oxide thickness, location and recess amount, and hence is robust against process-induced variations. Thus, the iFinFET is truly an intriguing new candidate transistor structure for future CMOS technologies.

To mitigate the additional process complexity and significant associated cost with multiple-patterning techniques, a novel patterning method via tilted ion implantation (TII) [6.7] is discussed in this work. By leveraging etch rate enhancement via ion implantation of a thin masking layer [6.8], TII-enhanced patterning is capable of achieving sub-lithographic features and/or doubling the density of features, one that is capable of achieving arbitrarily small feature size, self-aligned to pre-existing features on the surface. A TII-based pitch-halving process flow is also described. TII-enhanced patterning is experimentally validated to be able to pattern features with dimensions below 10 nm and with lower line-edge roughness than that of pre-existing masking features on the surface

of a substrate, in a self-aligned manner. Since ion implantation is a well-established technique with good process control and high throughput, the TII double-patterning approach can be easily adopted in high-volume manufacturing processes and shows promise for extending IC technology advancement beyond the 7 nm technology node (sub-40 nm pitch).

The semiconductor industry has identified increasing cost as the focus of concerns and doubts over the vitality of Moore's Law going forward [6.9]. This work conducts a survey regarding IC manufacturing cost throughout the industry. Then case studies reveal that the iFinFET technology and TII double patterning technique in this work have significant economic merit in future technology nodes, especially beyond the 7 nm technology node where the industry does not yet have clear solutions.

6.2 Suggestions for Future Work

The iFinFET work described in **Chapter 3** proposes an intriguing new candidate transistor structure for future CMOS technology scaling. To validate the suitability of replacing the conventional bulk FinFET with the iFinFET in advanced technology nodes, it is desirable to experimentally demonstrate the performance and scalability benefits of the iFinFET in state-of-the-art technologies. For the first experiment, a single inserted-oxide iFinFET can be fabricated using a conventional SOI substrate (cf. **Fig. 3.16**).

Also, the iFinFET structure facilitates a novel 6-T SRAM cell implementation. Unlike adjusting the number of fins in the pull-up (PU), pull-down (PD) and pass-gate (PG) devices as is done in the FinFET SRAM cell, the iFinFET SRAM cell consists only 1 "fin" in the PU, PD and PG devices. This will save a lot of chip area. The alpha and beta ratios of iFinFET-based SRAM cell can be tuned by adjusting the number of Si channels each device has. This is straightforward to implement in reality, because SiO_2 serves as an ideal etch stop layer for Si.

Furthermore, as is mentioned in Chapter 3, the iFinFET structure is compatible with implementing multiple gate-oxide thicknesses as needed for system-on-chip (SoC) applications. Thus, it is also worthwhile to investigate the iFinFET's behavior as an I/O device.

The idea of TII patterning in **Chapter 4** is fundamentally based on the etch rate enhancement via ion implantation. In reality, the etch rate of the making layer can be either enhanced or retarded in the implanted region, depending on what material is used as the masking layer. In other words, TII patterning can also be achieve in an etch-rate-retarded fashion. Also, different ion species may be explored along with optimizing the substrate temperature during implantation to achieve the patterning resolution in TII patterning.

Finally, TII patterning can be implemented broadly for any application that requires a small feature size, such as a short gate length in transistors or a small contact gap in micro/nanoelectromechanical system (M/NEMS) devices.

6.3 References

[6.1] G. E. Moore, “Cramming more components onto integrated circuits”, *Electronics*, volume 38, number 8, April 19, 1965, pp.114-117.
[6.2] C. Auth, C. Allen, A Blattner, D. Bergstrom, M. Brazier, M. Bost, M. Buehler, V. Chikarmane, T. Ghani, T. Glassman, R. Grover, W. Han, D. Hanken, M. Hattendorf, P. Hentges, R. Heussner, J. Hicks, D. Ingerly, P. Jain, S. Jaloviar, R. James, D. Jones, J. Jopling, S. Joshi, C. Kenyon, H. Liu, R. McFadden, B. McIntyre, J. Neirynck, C. Parker, L. Pipes, I. Post, S. Pradhan, M. Prince, S. Ramey, T. Reynolds, J. Roesler, J. Sandford, J. Seiple, P. Smith, C. Thomas, D. Towner, T. Troeger, C. Weber, P. Yashar, K. Zawadzki, and K. Mistry, “A 22nm high performance and low-power CMOS technology featuring fully-depleted tri-gate transistors, self-aligned contacts and high density MIM capacitors,” *VLSI Technol. (VLSIT), 2012 Symp*., pp. 131–132, 2012.
[6.3] International Technology Roadmap for Semiconductors (ITRS), 2011 (http://public.itrs.net/)
[6.4] P. Zheng, Y.-B. Liao, N. Damrongplasit, M.-H. Chiang, T.-J. King Liu, “Variation-Aware Comparative Study of 10-nm GAA Versus FinFET 6-T SRAM Performance and Yield, ” *IEEE Transactions on Electron Devices,* vol 61, no. 12, pp. 3949 – 3954, 2014.
[6.5] P. Zheng, D. Connelly, F. Ding, T.-J. King Liu, “Simulation-Based Study of the Inserted-Oxide FinFET for Future Low-Power System-on-Chip Applications,” *IEEE Electron Device Letters*, vol. 36, no. 8, pp. 742-744, 2015.
[6.6] P. Zheng, D. Connelly, F. Ding, T.-J. King Liu "FinFET Evolution Toward Stacked-Nanowire FET for CMOS Technology Scaling", *IEEE Transactions on Electron Devices,* vol 62, no. 12, pp. 3945 – 3950, 2015.
[6.7] S. W. Kim, P. Zheng, K. Kato, L. Rubin, T.-J. King Liu, “Enhanced Patterning by Tilted Ion Implantation’” *Proc. SPIE Alternative Lithographic Technologies VIII,* 2016, 9777, 97771B1-97771B6.
[6.8] X. Sun, Q. Lu, H. Takeuchi, S. Balasubramanian, T.-J. King Liu, “Selective Enhancement of SiO_2 Etch Rate by Ar Ion Implantation for Improved Etch Depth Control,” *Electrochemical and Solid-State Letters* 2007, 10(9), pp. 89-91.
[6.9] W. M. Holt, “Moore’s Law: A Path Forward”, *2016 International Solid State Circuits Conference*, paper 1.1.