Advanced Source/Drain Technologies for Nanoscale CMOS

by

Pankaj Kalra

B. Tech. (Banaras Hindu University) 2003

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 Chenming Hu
Professor Eugene E. Haller

Fall 2008

The dissertation of Pankaj Kalra is approved:

Professor Tsu-Jae King Liu, Chair — Date

Professor Chenming Hu — Date

Professor Eugene E. Haller — Date

University of California, Berkeley

Fall 2008

Advanced Source/Drain Technologies for Nanoscale CMOS

Copyright © 2008

by

Pankaj Kalra

Abstract

Advanced Source/Drain Technologies for Nanoscale CMOS

by

Pankaj Kalra

Doctor of Philosophy in Engineering – Electrical Engineering and Computer Sciences

University of California, Berkeley

Professor Tsu-Jae King Liu, Chair

Transistor scaling has been the driving force for technology advancements in the semiconductor industry over the last few decades. In order to mitigate short channel effects, the gate-oxide thickness and source/drain junction depth have been scaled along with the gate length. Recently, however, gate-oxide thickness scaling has slowed, as evidenced by the fact that an equivalent oxide thickness (EOT) of ~1 nm has been used for the past 2-3 generations of CMOS technology. Although significant progress has been made in the development of high-permittivity (high-κ) gate-dielectric materials and metal gate technology in recent years, it will be difficult to scale EOT well below 1 nm. This makes junction-depth scaling even more pressing for continued transistor scaling. Furthermore, as the dimensions of MOSFETs are scaled down, the contact resistance of silicide-to-source/drain regions increasingly limits transistor performance. This is because the on-state resistance of a MOSFET drops with transistor scaling, whereas contact resistance increases with contact area scaling. Contact resistance increases exponentially

with Schottky barrier height (SBH) of the silicide-to-semiconductor contact. Thus, lower values of SBH will be needed in order to achieve substantial performance improvements with transistor scaling in the future. In practice, fermi-level pinning makes it especially difficult to attain low values of SBH for metal (silicide) contact to *n*-type silicon. This dissertation addresses the aforementioned scaling challenges associated with the design of source/drain structures for sub-45 nm CMOS generations.

Firstly, the progress made towards the formation of ultra-shallow junctions with the help of advanced annealing techniques, low-energy implants, and GCIB doping is presented. The experimental results obtained with flash annealing indicate that it is possible to achieve sub-15 nm junctions with lower sheet resistance (~1000 Ω/), adequate for 32 nm CMOS technology.

Since high-κ/metal-gate stacks are already used in the most advanced 45 nm CMOS technology today, it is important to assess the compatibility of flash annealing with high-κ/metal-gate stacks. The process integration of high-κ/metal-gate stacks with flash annealing is discussed next. It is shown that the flash annealing process has minimal effects on gate stack properties and is found to be compatible with the high-κ/metal-gate stacks. However, it results in degraded interface quality which is improved by using a post-metallization anneal.

To reduce the effective SBH of silicide-to-semiconductor contact, various species (nitrogen, fluorine, sulfur and selenium) are studied. These species were implanted into the semiconductor, and then "piled up" at the silicide-semiconductor interface during the silicidation process. It is shown that significant SBH lowering (by as much as 0.37 eV) can be achieved on *n*-type silicon using nitrogen. The impact of this process on the

properties of NiSi is assessed and the mechanism of SBH reduction is explained. Encouraging results are also obtained with sulfur and selenium, and a comparison of effective SBH reduction is made for all studied species.

Finally, material properties of nickel germanide formed on epi-Ge on Si substrate are studied to form low-resistance and thermally stable contact material for realizing highly-scaled high-performance technology based on Ge-channel MOSFETs.

____

Professor Tsu-Jae King Liu, Chair

To my parents, for their love, support and encouragement

Table of Contents

Chapter 1: Introduction **1**

1.1. MOSFET Scaling Benefits 1

1.2. Bulk Si-MOSFET Scaling Challenges 3

1.3. Emerging CMOS Technology Roadmap 5

1.4. Challenges of Source/Drain Design for Nanoscale CMOS 8

1.5. Dissertation Objectives and Outline 13

1.6. References 14

Chapter 2: Formation of Ultra-Shallow Junctions (USJs) **20**

2.1. Introduction 20

2.2. State-of-the-art Junction Technology 25

2.3. Alternate Doping and Annealing Techniques 28

2.4. Junction Metrology 32

2.5. Experimental Results with Spike Anneal 36

2.6. Experimental Results with Laser Anneal 40

2.7. Experimental Results with Flash Anneal 42

2.8. Experimental Results with Infusion (GCIB) Doping 54

2.9. Summary 59

2.10. References 60

Chapter 3: Process Integration of High-*k*/Metal-Gate with Flash Anneal...............**66**

3.1. Introduction..66

3.2. Advanced Gate Stacks 68

3.3. Device Fabrication 71

3.4. Device Characterization 72

3.4.1. SCE Control...72

3.4.2. Gate Stack Integrity 74

3.4.3. MOSFET Performance 81

3.4.4. Effects of Post-Metallization Anneal 87

3.4.5. Challenges with Flash Anneal 89

3.5. Summary...91

3.6. References 92

Chapter 4: Low Barrier Height Contact Technology **101**

4.1. Introduction..101

4.2. Approaches to Reduce Barrier Height...107

4.2.1. Bandgap Engineering 108

4.2.2. Dual Silicide Technology 110

4.2.3. Interface Passivation..110

4.2.4. Image-force Induced Barrier Lowering 112

4.2.5. Workfunction Engineering..113

4.3. Experimental Details 115

4.4. Results and Discussion 117

4.4.1. Effects of Nitrogen Implantation 118

4.4.2. Effects of Fluorine Implantation 126

4.4.3. Effects of Selenium and Sulfur Implantation 128

4.5. Summary 131

4.6. References 132

Chapter 5: Contact Material for Ge MOSFETs 137

5.1. Introduction 137

5.2. Heteroepitaxial Growth of Ge on Silicon 140

5.3. Metal Germanides 142

5.4. Experimental Details 143

5.5. Physical Analysis 145

5.6. Wet Chemical Reaction 147

5.7. Summary 149

5.8. References 150

Chapter 6: Conclusions 154

6.1. Summary 154

6.2. Contributions 155

6.3. Recommendations for Future Research 159

6.4. References 160

Acknowledgements

I am very grateful to so many people for their help and support given to me along the entire course of my graduate study. Without their help, this thesis would not have been possible. First of all, I would like to express my sincere gratitude to my advisor Professor Tsu-Jae King Liu, for her inspiring guidance and constant encouragement through out this work. Her vast knowledge and strong technical expertise have helped me in getting an excellent background in the semiconductor device technology, identifying the opportunities in this field and preparing for the challenges. I am thankful for the rewarding learning experience that her mentorship has produced.

I would like to thank Professor Chenming Hu and Professor Eugene Haller for serving on my dissertation committee. I would also like to thank Professor Nathan Cheung for being on my qualifying examination committee.

Special thanks go to Dr. Hsing-Huang Tseng and Dr. Prashant Majhi at SEMATECH, Austin for providing me the opportunity to be part of their team and to gain valuable industrial experience. In particular, I would like to thank Prashant for being an excellent mentor during my internship, and for his support and friendship. I would also like to thank Dr. Raj Jammy, Dr. Muhammad Mustafa Hussain, Dr. Rusty Harris, Nikhil Vora, Dr. P. Y. Hung, Dr. Dawei Heh, Dr. Gennadi Bersuker, Dr. Chadwin Young, Dr. Jungwoo Oh, Dr. Huang-Chun Wen, Joel Barnett and Barry Sassman.

Additionally, I benefited greatly from the fruitful research collaborations with Dr. Michael Current (Frontier Semiconductor) for non-contact sheet resistance measurements, Steve McCoy (Mattson Technology) for flash annealing, Dr. John Hautala (Epion-TEL)

for GCIB doping and Professor Hyunsang Hwang (GIST, Korea) for post-metallization anneal. Their contributions became the essential part of various research projects.

I would like to express my warmest appreciation for Hideki Takeuchi who generously helped me with various fabrication tools in Microlab. His invaluable experience and critical feedback helped me to gain better insight into the nuances of the nanoscale device fabrication. I would also like to acknowledge the following current or former members of device group that have helped me to get through the initial learning curve as a new graduate student and microlab user: Sriram Balasubramanian, Joanna Lai, Vidya Vardarajan, Mohan Dunga, Hiu Yung Wong, Carrie Low and Daewon Ha. I must thank the visiting fellows to the Device Group, Akira Hokazono, Takuro Matsutoya, Yuri Yasuda, Koichi Fukuda, Taro Osabe, Si-Woo Lee and Woo Young Choi for sharing their experiences. The administrative and technical staffs at Microlab and ATDF, Austin deserve sincere thanks for their hard work to keep the fabrication facilities functional.

I would also like to thank the wonderful EE Graduate Student Affairs staff, Ruth Gjerde and Mary Byrnes for giving countless hours of service to the graduate students. Thanks to Linda Manly, Charlotte Jones, Misty Kiuchi and Sim Kallan for being very helpful with monetary matters.

Much needed financial support by MARCO Materials, Structures and Devices (MSD) Center, UC Discovery Grant, Intel, Infineon, Applied Materials Fellowship and SEMATECH is gratefully acknowledged.

My heartfelt thanks to my friends Akira, Alvaro, Anshul, Arka, Bala, Blake, Chung-Hsun, Darsen, Donovan, Drew, Dunga, Gaurav, Gautam, Huang-Chun, Joanna, Kaushik, Kinyip, Koichi, Krish, Marie, Nikhil, Noel, Rahul, Sriram, Steve, Takuro, Taro,

Vidya, Vishnu, Woo Young, Yung, Yuri and many others for their friendship and support.

Mere words are not enough to express the gratitude I feel for my family. I cherish our times together.

Chapter 1

Introduction

1.1 MOSFET Scaling Benefits

First stated in 1965, Moore's law describes the unparalleled technology advancement over the past 40 years which has allowed the number of transistors on a chip to double about every two years [1]. This phenomenal progress has been made possible by continual downscaling of the metal-oxide-semiconductor field-effect transistor (MOSFET) to smaller physical dimensions. MOSFETs have the remarkable feature that as they become smaller they also become cheaper, consume less power, become faster, and enable more functions per unit area of silicon. As a result, denser silicon integrated circuits (ICs) can be realized, offering superior performance at reduced cost per function as shown in Figure 1.1 and 1.2.

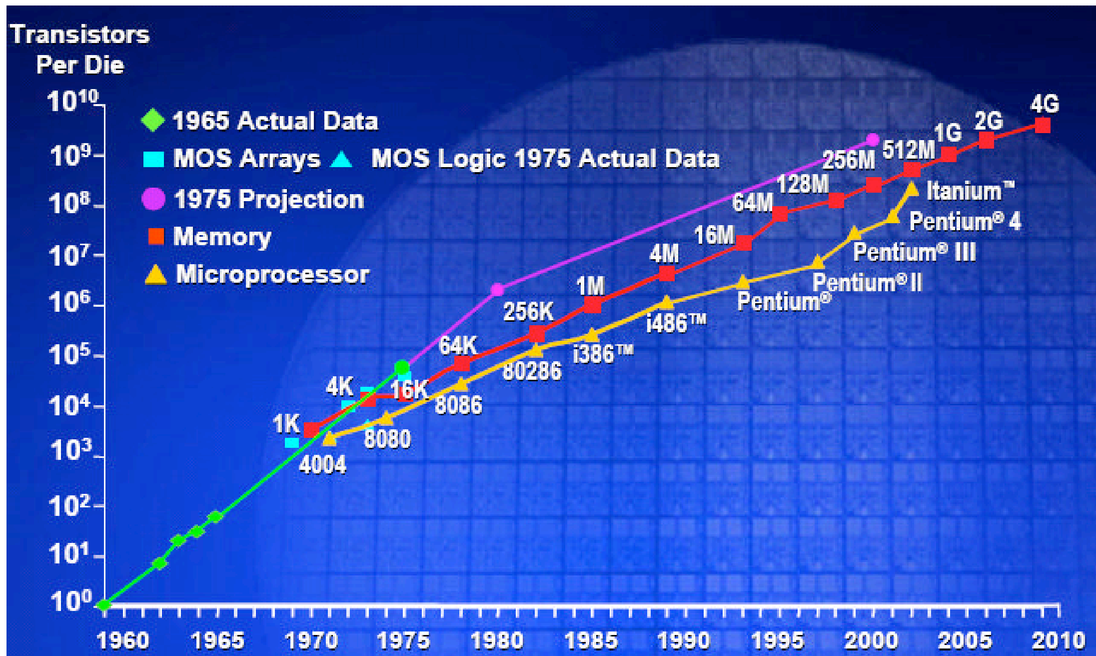


Figure 1.1. Number of transistors on a chip, as a function of the first year of production [2].

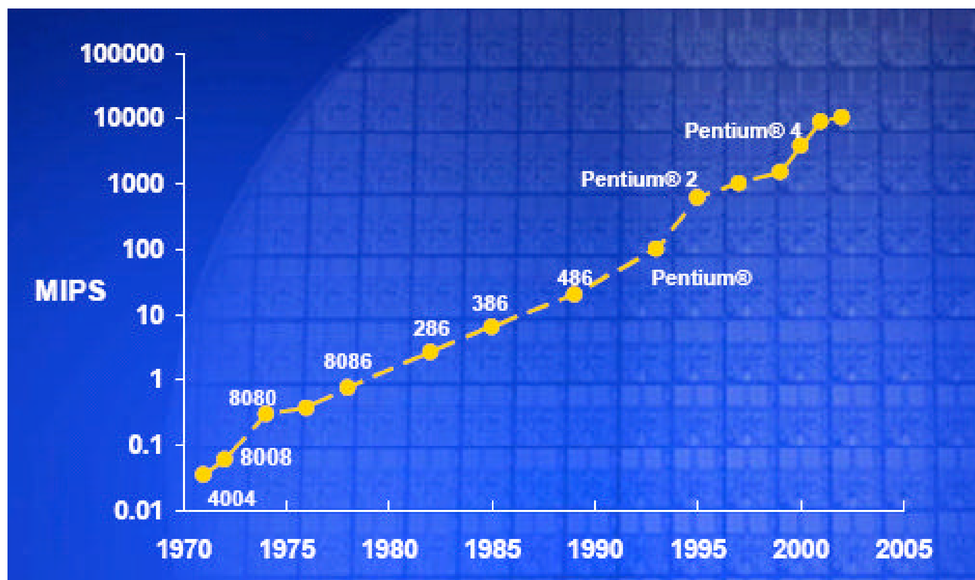


Figure 1.2. The increase in chip performance, measured in millions of instructions per second (MIPS), as a consequence of scaling [2].

1.2 Bulk Si-MOSFET Scaling Challenges

Figure 1.3 shows the schematic of the conventional planar bulk Si-MOSFET, the basic building block of an integrated circuit. As the gate length (L_g) of a MOSFET decreases, the capacitive coupling of the channel potential to the source and the drain terminals increases with respect to the capacitive coupling to the gate terminal, resulting in short channel effects (SCE), such as threshold voltage (V_{TH}) roll-off? smaller V_{TH} at shorter L_g, and drain-induced barrier lowering (DIBL)? smaller V_{TH} at higher drain voltage (V_{DS}) due to reduction of the source-channel potential barrier by the drain voltage. V_{TH} roll-off and DIBL manifest into increased off-state transistor leakage (I_{off}). I_{off} is becoming a severe concern for high-performance logic technologies that may hinder CMOS scaling because of significant passive power consumption as shown in Figure 1.4 [2, 3].

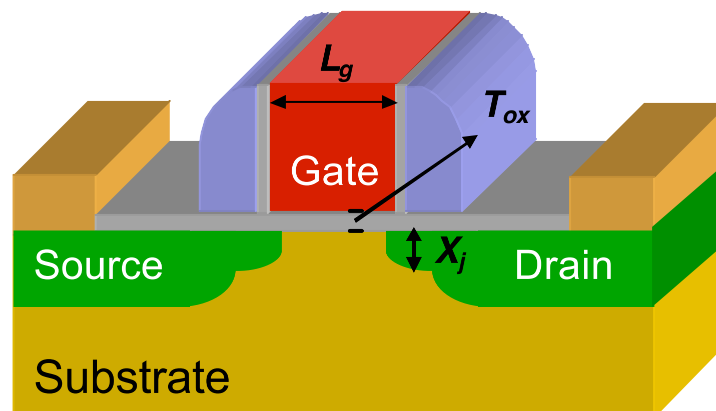


Figure 1.3. Schematic diagram of the bulk Si MOSFET.

Therefore, careful device design is required so that SCE do not prevent the use of a minimum L_g MOSFET. For bulk MOSFETs, a minimum L_g must be ~5l, where l is the characteristic length and is given by [4]:

$$l = 0.1\left(X_j T_{ox} T_{dep}^2\right) \tag{1.1}$$

where X_j is the source/drain extension junction depth, T_{ox} is the gate dielectric (SiO_2) thickness and T_{dep} is the channel depletion depth. According to Eq. (1.1), vertical dimensions (T_{ox}, X_j, T_{dep}) must be scaled together with L_g to suppress SCE in bulk MOSFETs.

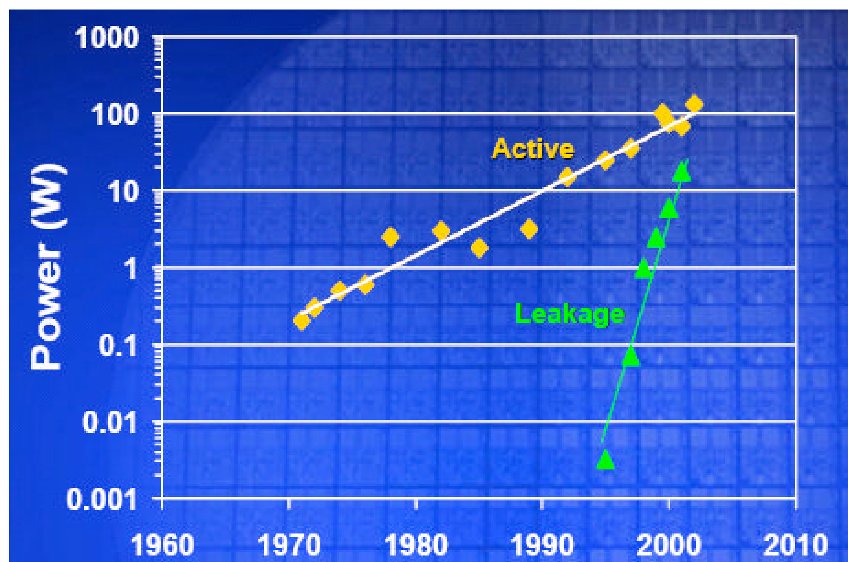


Figure 1.4. The trend in active and leakage power of microprocessors as a consequence of scaling [2]. The leakage power is approaching ~50% of the total power in a modern-day microprocessor.

This scaling methodology has worked very well for several decades, but as silicon CMOS technology advances into the nanometer regime, fundamental and practical limits impede the traditional scaling of transistors as discussed next. Historically, the gate dielectric thickness has been the single most important dimension to enable device scaling. Scaling down of the gate dielectric not only increases capacitive coupling of the gate to the channel but also leads to increased on-state transistor drive current (I_{on}).

However, the thickness of SiO_2-based gate dielectrics is approaching physical limits (<2nm) resulting in severe gate leakage current due to quantum mechanical tunneling of carriers across the thin dielectric. To scale down T_{dep}, higher channel doping density is necessary. Reducing T_{dep} helps to eliminate leakage paths far from the dielectric/channel interface. In sub-100 nm MOSFETs, a halo implant is generally used to suppress sub-surface leakage, but this increases the average doping in the channel [5]. The increase in channel doping leads to degraded carrier mobility due to higher vertical electric field and more impurity scattering. For very high channel doping near the source/drain extensions, band-to-band tunneling also becomes an issue. Furthermore, as the channel volume shrinks in ultra-scaled transistors, the statistical fluctuation of dopant atoms causes device-to-device variations [6, 7]. Shallow X_j reduces the source/drain-channel coupling relative to the gate-channel coupling. However, the reduced junction depth increases parasitic source/drain resistance, resulting in degraded on-state transistor drive current.

1.3 Emerging CMOS Technology Roadmap

In order to maintain the historical 17% per year performance improvement rate, various novel processes and materials have been introduced to the bulk Si-MOSFET structure. Figure 1.5 shows the insertion of the new materials for Intel's logic technologies [8].

In the most recent technology nodes, much effort has been directed towards increasing the carrier mobility (and thus improving I_{on}) by strain engineering. However, the wafer-level strain methods such as tensile-stressed Si on relaxed $Si_{1-x}Ge_x$ [9-12] and compressively-stressed $Si_{1-x}Ge_x$ on Si [13, 14] did not make their way to high-volume production. On the other hand, uniaxial process-induced strain has been introduced into

large-scale manufacturing [15-17] and is being adopted in all sub-100 nm high-performance logic technologies (Figure 1.6).

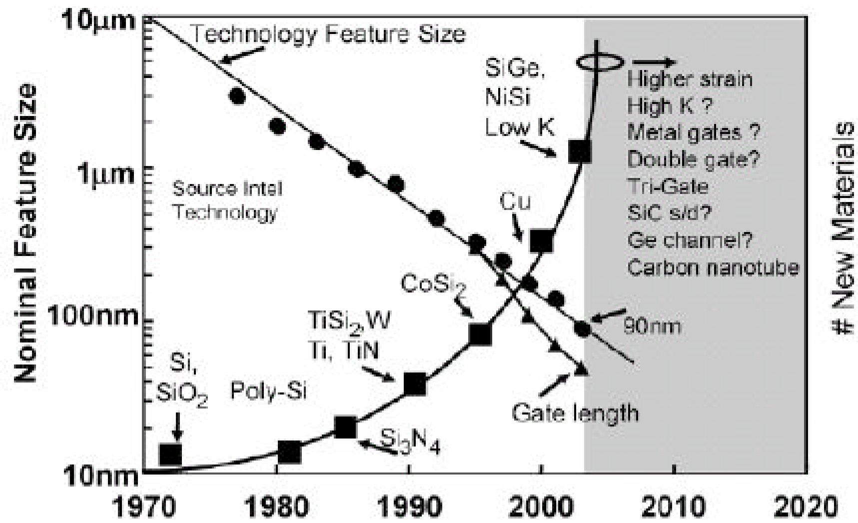


Figure 1.5. Feature size and new materials versus year of production [8].

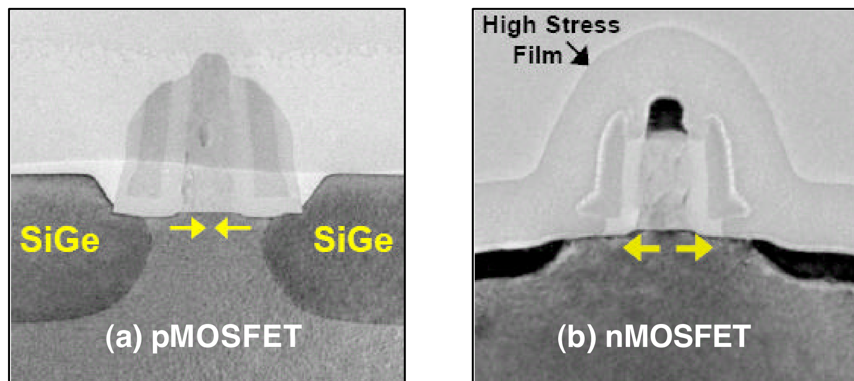


Figure 1.6. Intel's 90nm logic technology [17]? (a) For pMOSFET, strained epitaxial SiGe film is embedded into the source/drain to induce compressive channel strain, (b) nMOSFET is capped with silicon nitride stressor to induce tensile strain in the channel.

Intense efforts have been applied toward finding an alternative gate dielectric material with higher permittivity (high-κ) to continue T_{ox} at low gate leakage. For the first

time, transistors with high-κ dielectrics and metal gate electrodes have been integrated into a manufacturable 45 nm process (Figure 1.7) [18]. The introduction of the metal gate electrodes brings the additional benefit of eliminating the poly-Si depletion effect.

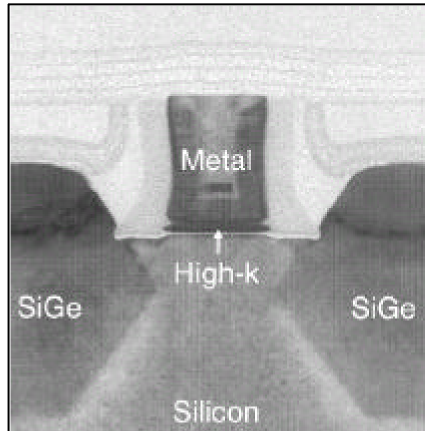


Figure 1.7. High-κ/metal-gate stack used in 45 nm high-performance logic technology [18].

Scaling CMOS to and beyond the 22 nm node will probably require structural changes to the planar silicon transistor and/or use of high-mobility channel material. The advanced transistor structures such as ultra-thin body FET (UTBFET) and multi-gate FET (MuGFET) are seen as potential replacements for the bulk MOSFET because of their inherent superior electrostatic integrity [19-21]. Another promising alternative is the adoption of germanium (Ge) as the channel material. *p*-channel Ge-MOSFETs have been proposed for highly-scaled high-performance technology due to higher bulk mobility of holes in Ge [22-25].

This section has provided a brief introduction to the most recent and possible technologies in the future to extend CMOS scaling. However, as will be seen in the next section, the source/drain series resistance will severely limit MOSFET performance as

CMOS is scaled deeper into the nano-scale regime irrespective of device architecture (bulk, UTBFET or MuGFET).

1.4 Challenges of Source/Drain Design for Nanoscale CMOS

As device dimensions are scaled down, the on-resistance ($R_{on}=V_{dd}/I_{on}$) of the intrinsic device is improved (reduced) and it becomes important to limit the source/drain series resistance (R_{series}) to be a small fraction of R_{on} to satisfy the device performance requirements (Figure 1.8). Historically, R_{series} has been kept to ~10% of R_{on} as shown in Figure 1.9 [26]. While device performance has increased for each technology node (more recently by reducing R_{on} via strain engineering), it has been difficult to scale down R_{series} for sub-100nm generations. Figure 1.10 shows the projected maximum R_{series} values *vs.* year of production in the near future that allow for the required I_{on} as set by *International Technology Roadmap for Semiconductors* [27]. For technology year >2008, R_{series} is expected to be ~20% of the total R_{on}, a significant fraction, yet no known solutions are

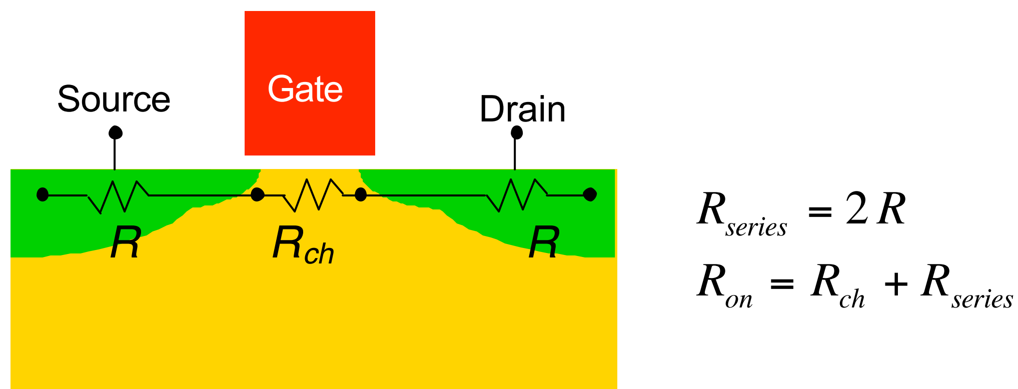


Figure 1.8. Schematic representation of the channel resistance (R_{ch}) and the parasitic S/D series resistance (R_{series}). The total on-resistance (R_{on}) of the MOSFET is the summation of R_{ch} and R_{series}.

available to meet or limit these projections. Therefore, R_{series} reduction presents one of the biggest challenges for continued aggressive CMOS scaling.

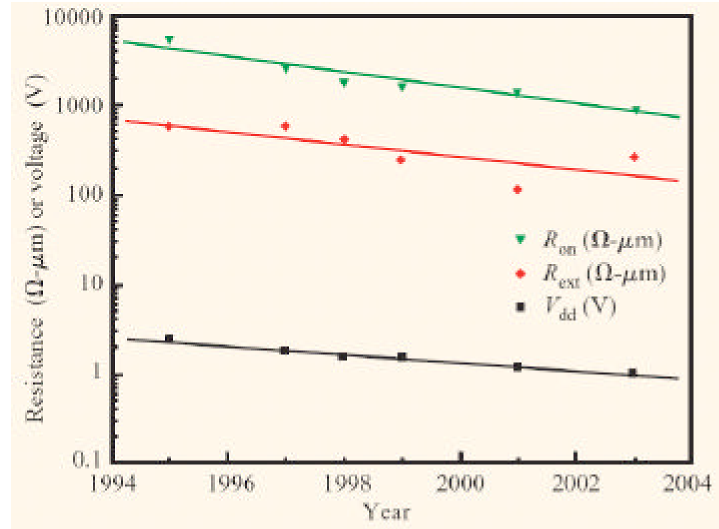


Figure 1.9. Historic scaling trends for R_{on} and R_{series} (or R_{ext}) of high-performance nMOSFETs [26].

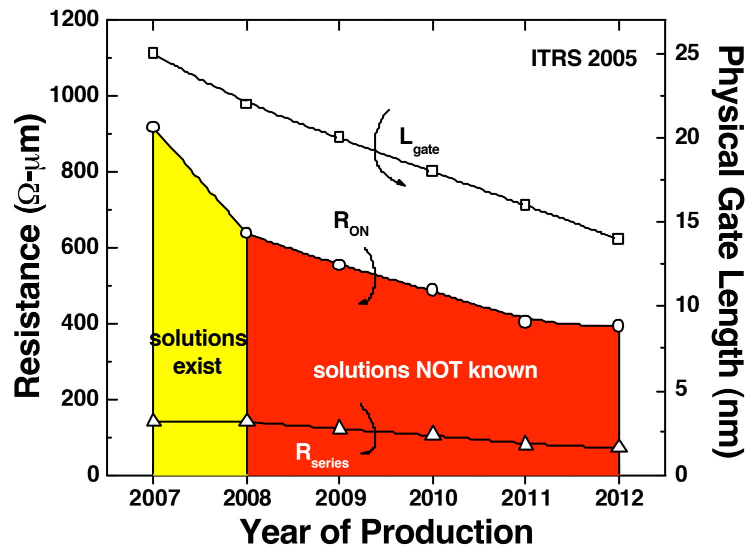


Figure 1.10. Scaling projections of R_{on} and R_{series} for high-performance bulk nMOSFETs [27]. R_{on} is calculated using the projected values of drive current and power supply so that the 17% per year device performance improvement targets are met for a given technology year. The red region indicates that the solutions to achieve these projections are not known.

As shown in Figure 1.11, R_{series} can be divided into the four components: (i) extension-to-gate overlap resistance (R_{ov}); (ii) S/D extension resistance (R_{ext}); (iii) deep S/D resistance ($R_{S/D}$); and (iv) contact resistance at the silicide-silicon interface (R_{co}). A simple first order estimation of these components is presented next to understand their dependence on device and process parameters.

The overlap resistance (R_{ov}) is a strong function of doping concentration in the overlap region and the lateral abruptness of the overlap doping profile. These two factors govern the accumulation carrier density and the current spreading. Increased lateral abruptness is needed for improved SCE, reduced spreading resistance and reduced accumulation resistance.

$$R_{ov} = f\left(Abruptness, N_{ext}, R_{ch}\right) \tag{1.2}$$

where N_{ext} is the peak dopant concentration in the extension region and R_{ch} is the channel resistance.

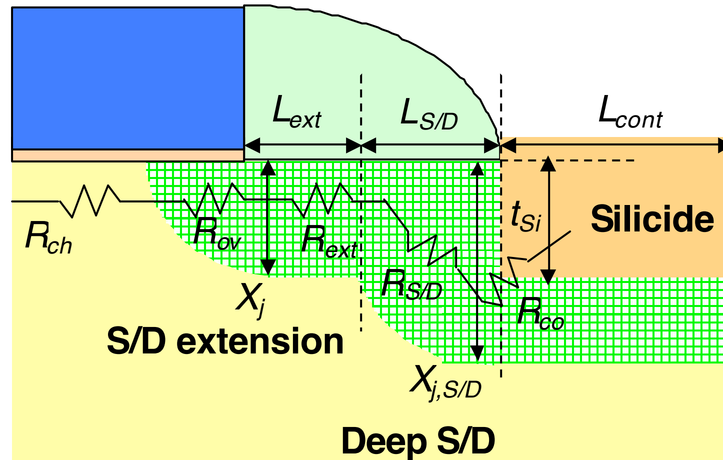


Figure 1.11. Schematic representation of various components of S/D series resistance.

Assuming an ideal box-like profile, the S/D extension resistance (R_{ext}) can be written as:

$$R_{ext} = \frac{\rho_{ext} L_{ext}}{W X_j} \quad (1.3)$$

where ρ_{ext} is the resistivity of the extension region, L_{ext} is the extension length, W is the channel width and X_j is its junction depth. Similarly, the deep S/D resistance ($R_{S/D}$) equals:

$$R_{S/D} = \frac{\rho_{S/D} L_{S/D}}{W\left(X_{j,S/D} - t_{Si}\right)} \quad (1.4)$$

where $\rho_{S/D}$ is the resistivity of the deep S/D region, $L_{S/D}$ is the lateral diffused length of deep junction, $X_{j,S/D}$ is its junction depth and t_{Si} is the thickness of silicon consumed during silicidation. The summation of R_{ext} and $R_{S/D}$ is commonly referred as the spreading resistance under the sidewall spacer (R_{spr}).

The contact resistance (R_{co}) is defined as the resistance between the silicide and the Si underneath the silicide and is expressed as:

$$R_{co} = \frac{\rho_c}{W L_{cont}} \quad (1.5)$$

where ρ_c is the specific contact resistivity of the silicide-silicon interface and is determined by the active dopant concentration and Schottky barrier height at the interface. L_{cont} is the length of the silicide region. R_{co} has another sub-component due to an additional current path at the sidewall of recessed contact region, which is a strong function of the silicon recess (t_{Si}).

For successful optimization of the source/drain design for lower R_{series}, it is important to first understand the relative contribution of its components. Figure 1.12 shows the results of an advanced series resistance model for sub-100 nm bulk-Si

technologies [28, 29]. The series resistance trend versus technology scaling predicts that R_{spr} and R_{co} are the dominant components and account for >85% of the total series resistance for all technologies studied. Furthermore, the contribution of R_{co} increases as the technology scales. This is not unexpected because of the requirement of shallower junctions for SCE control and the scaling of L_{cont}, which shrinks by a factor of ~0.7× every node. Shallow junction depth and smaller L_{cont} degrades R_{ext} and R_{co}, respectively. A similar trend is seen for 90 nm SOI technology [30]. The extracted series resistance breakdown indicates that the contribution of R_{spr} and R_{co} is ~95% as seen in Figure 1.13. It has been reported that the contact resistance between the silicide and Si-fin will be a serious challenge to MuGFET structures as well [31, 32].

To summarize, the source/drain design for future nanoscale CMOS devices requires reduced R_{ext} and R_{co}. However, ultra-shallow junctions are a key requirement to suppress short channel effects, but this leads to higher R_{ext}. Therefore, very high active

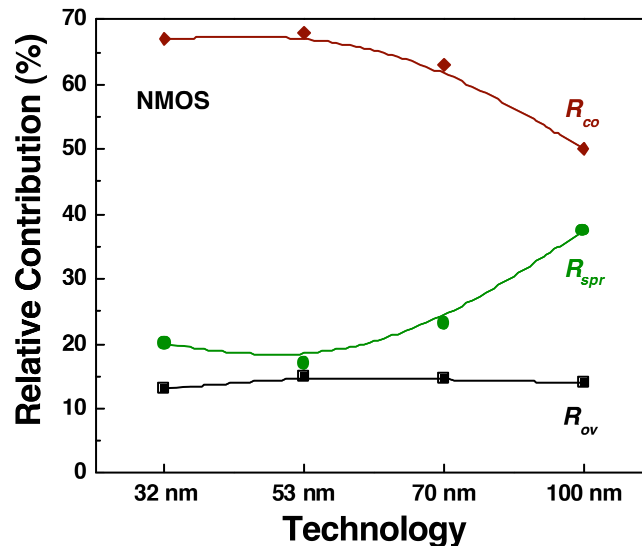


Figure 1.12. The relative contribution of the various components of the series resistance for different technology nodes. The device parameters for each technology are scaled down according to *ITRS* (Source: Prof. Jason Woo, UCLA).

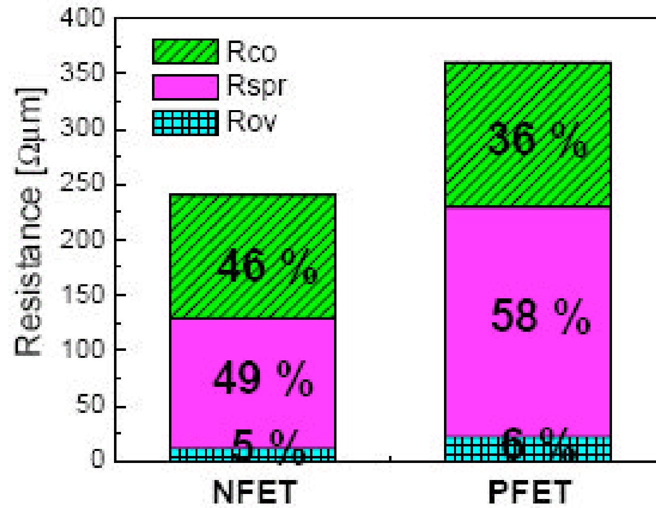


Figure 1.13. The extracted series resistance and the contributions of its components for 90 nm SOI CMOS technology [30].

dopant concentration levels and abruptness are required for these junctions. To limit the impact of R_{co}, lower ρ_c is desirable. Since ρ_c decreases exponentially with decreasing Schottky barrier height (SBH) of the silicide-Si contact, a lower SBH will be needed.

1.5 Dissertation Objectives and Outline

In this dissertation, some of the aforementioned source/drain engineering issues are addressed, with the aim of minimizing the impact of parasitic series resistance on device performance. The process solutions presented here are applicable not only to the bulk-Si MOSFET but also to advanced transistor structures.

A systematic study to meet the stringent junction requirements for future CMOS generations using conventional and/or advanced processing methods is presented in Chapter 2.

In Chapter 3, the compatibility of millisecond flash annealing with advanced gate stacks in a "gate-first" CMOS process flow is addressed. The evaluation of the flash

annealed devices is presented to understand the impact of flash annealing on high-κ/metal-gate device performance and reliability.

Chapter 4 focuses on reducing the Schottky barrier height of a silicide to *n*-type Si contact by introducing impurities into the silicon via ion implantation. These impurities are then piled up at the silicide-semiconductor interface during silicidation. Various impurity species are studied using this process, and comprehensive electrical and physical analyses are performed on the bulk silicide and silicide-Si interface.

Chapter 5 presents a brief introduction to epi-Ge MOSFETs. To fully exploit the superior carrier transport properties of Ge, low-resistance, thermally-stable, and self-aligned contact material with low formation temperature is needed. Nickel germanide is investigated as the choice of germanide for source/drain contact in epi-Ge MOSFETs.

Chapter 6 summarizes the key contributions of this research and provides suggestions for future research directions.

1.6 References

[1] G. E. Moore, "Cramming more components onto integrated circuits," *Electronics*, vol. 38, pp. 114 -117, 1965.

[2] G. E. Moore, "No Exponential is Forever: But "Forever" Can be Delayed!," *IEEE ISSCC Tech. Digest*, pp. 20-23, 2003.

[3] W. Haensch, E. J. Nowak, R. H. Dennard, P. M. Solomon, A. Bryant, O. H. Dokumaci, A. Kumar, X. Wang, J. B. Johnson, and M. V. Fischetti, "Silicon CMOS Devices Beyond Scaling," *IBM Journal of Research and Development*, vol. 50, no. 4/5, pp. 339-361, 2006.

[4] Z. H. Liu, C. Hu, J. H. Huang, T. Y. Chan, M. C. Jeng, P. K. Ko, and Y. C. Cheng, "Threshold Voltage Model for Deep-submicrometer MOSFETs," *IEEE Transactions on Electron Devices*, vol. 40, pp. 86-95, 1993.

[5] Y. Taur, C. H. Wann, and D. J. Frank, "25 nm CMOS Design Considerations," *IEEE IEDM Tech. Digest*, pp. 789-792, 1998.

[6] A. Asenov, "Random Dopant Induced Threshold Voltage Lowering and Fluctuations in Sub-0.1 μm MOSFET's: A 3-D "Atomistic" Simulation Study," *IEEE Transactions on Electron Devices*, vol. 45, pp. 2505-2513, 1998.

[7] T. Tanaka, T. Usuki, T. Futatsugi, Y. Momiyama, and T. Sugii, "V_{th} Fluctuation Induced by Statistical Variation of Pocket Dopant Profile," *IEEE IEDM Tech. Digest*, pp. 271-274, 2000.

[8] S. E. Thompson, R. S. Chau, T. Ghani, K. Mistry, S. Tyagi, and M. T. Bohr, "In Search of "Forever" Continued Transistor Scaling One New Material at a Time," *IEEE Transactions on Semiconductor Manufacturing*, vol. 18, no. 1, pp. 26-36, Feb 2005.

[9] J. Welser, J. L. Hoyt, and J. F. Gibbons, "NMOS and PMOS Transistors Fabricated in Strained Silicon/Relaxed Silicon-germanium Structures," *IEEE IEDM Tech. Digest*, pp. 1000-1002, 1992.

[10] K. Rim, J. L. Hoyt, and J. F. Gibbons, "Transconductance Enhancement in Deep Submicron strained Si n-MOSFETs," *IEEE IEDM Tech. Digest*, pp. 707-710, 1999.

[11] T. Tezuka, N. Sugiyama, T. Mizuno, and S. Takagi, " High-performance Strained Si-on-insulator MOSFETs by Novel Fabrication Processes Utilizing Ge-

condensation Technique," *IEEE VLSI Technology Symp. Tech. Digest*, pp. 96-97, 2002.

[12] K. Rim, J. Chu, K. A. Jenkins, T. Kanarshy, K. Lee, A. Mocuta, H. Zhu, R. Roy, J. Newbury, J. Ott, K. Petrarca, P. Mooney, D. Lacey, S. Koester, K. Chan, D. Boyd, M. Ieong, and H.-S. P. Wong, "Characteristics and Device Design of Sub-100 nm Strained-Si N- and P-MOSFETs," *IEEE VLSI Technology Symp. Tech. Digest*, pp. 98-99, 2002.

[13] C. W. Leitz, M. T. Currie, M. Lee, Z. Cheng, D. A. Antoniadis, and E. Fitzgerald, " Channel Engineering of SiGe-Based Heterostructures for High Mobility MOSFETs," *Mat. Res. Soc. Symp. Proc.*, vol. 686, pp. 113-118, 2002.

[14] M. Shima, T. Ueno, T. Kumise, H. Shido, Y. Sakuma, and S. Nakamura, "<100> Channel Strained-SiGe p-MOSFET with Enhanced Hole Mobility and Lower Parasitic Resistance," *IEEE VLSI Technology Symp. Tech. Digest*, pp. 94-95, 2002.

[15] S. Ito, H. Namba, K. Yamaguchi, T. Hirata, K. Ando, S. Koyama, S. Kuroki, N. Ikezawa, T. Suzuki, T. Saitoch, and T. Horiuchi, "Mechanical Stress Effect of Etch-stop Nitride and its Impact on Deep Submicron Transistor Design," *IEEE IEDM Tech. Digest*, pp. 247-250, 2000.

[16] A. Shimizu, K. Hachimin, N. Ohki, H. Ohta, M. Koguchi, Y. Nonaka, H. Sato, and F. Ootsuka, "Local Mechanical-Stress Control (LMC): A New Technique for CMOS Performance Enhancement," *IEEE IEDM Tech. Digest*, pp. 433-437, 2001.

[17] T. Ghani, M. Armstrong, C. Auth, M. Bost, P. Charvat, G. Glass, T. Hoffmann, K. Johnson, C. Kenyon, J. Klaus, B. McIntyre, K. Mistry, A Murthy, J. Sandford, M. Silberstein, S. Sivakumar, P. Smith, K. Zawadzki, S. Thompson, and M. Bohr, "A

90 nm High Volume Manufacturing Logic Technology Featuring Novel 45 nm Gate Length Strained Silicon CMOS Transistors," *IEEE IEDM Tech. Digest*, pp. 978–980, 2003.

[18] K. Mistry, C. Allen, C. Auth, B. Beattie, D. Bergstrom, M. Bost, M. Brazier, M. Buehler, A. Cappellani, R. Chau, C. -H. Choi, G. Ding, K. Fischer, T. Ghani, R. Grover, W. Han, D. Hanken, M. Hattendorf, J. He, J. Hicks, R. Huessner, D. Ingerly, P. Jain, R. James, L. Jong, S. Joshi, C. Kenyon, K. Kuhn, K. Lee, H. Liu, J. Maiz, B. Mclntyre, P. Moon, J. Neirynck, S. Pae, C. Parker, D. Parsons, C. Prasad, L. Pipes, M. Prince, P. Ranade, T. Reynolds, J. Sandford, L. Shifren, J. Sebastian, J. Seiple, D. Simon, S. Sivakumar, P. Smith, C. Thomas, T. Troeger, P. Vandervoorn, S. Williams, and K. Zawadzki, " A 45nm Logic Technology with High-k+Metal Gate Transistors, Strained Silicon, 9 Cu Interconnect Layers, 193nm Dry Patterning, and 100% Pb-free Packaging," *IEEE IEDM Tech. Digest*, pp. 247–250, 2007.

[19] Y. -K. Choi, K. Asano, N. Lindert, V. Subramanian, T. -J. King, J. Bokor, and C. Hu, "Ultrathin-body SOI MOSFET for Deep-sub-tenth Micron Era," *IEEE Electron Device Letters*, vol. 21, pp. 254-255, 2000.

[20] 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 Transactions on Electron Devices*, vol. 47, pp. 2320-2325, 2000.

[21] B. Doyle, B. Boyanov, S. Datta, M. Doczy, S. Hareland, B. Jin, J. Kavalieros, T. Linton, R. Rios, and R. Chau, "Tri-Gate Fully-depleted CMOS Transistors:

Fabrication, Design and Layout," *IEEE VLSI Technology Symp. Tech. Digest*, pp. 133-134, 2003.

[22] C. O. Chui, H. Kim, D. Chi, B. B. Triplett, P. C. McIntyre, and K. C. Saraswat, "A Sub-400°C Germanium MOSFET Technology with High-κ Dielectric and Metal Gate," *IEEE IEDM Tech. Digest*, pp. 437-440, 2002.

[23] C. H. Huang, M. Y. Yang, A. Chin, W. J. Chen, C. X. Zhu, B. J. Cho, M.-F. Li, and D. L. Kwong, "Very Low Defects and High Performance Ge-On-Insulator p-MOSFETs with Al_2O_3 Gate Dielectrics," *IEEE VLSI Technology Symp. Tech. Digest*, pp. 119-120, 2003.

[24] W. P. Bai, N. Lu, J. Liu, A. Ramirez, D. L. Kwong, D. Wristers, A. Ritenour, L. Lee, and D. Antoniadis, "Ge MOS Characteristics with CVD HfO_2 Gate Dielectrics and TaN Gate Electrode," *IEEE VLSI Technology Symp. Tech. Digest*, pp. 121-122, 2003.

[25] P. Zimmerman, G. Nicholas, B. De Jaeger, B. Kaczer, A. Stesmans, L.-A. Ragnarsson, D. P. Brunco, F. E. Leys, M. Caymax, G. Winderickx, K. Opsomer, M. Meuris, and M. M. Heyns, "High Performance Ge pMOS Devices Using a Si-Compatible Process Flow," *IEEE IEDM Tech. Digest*, pp. 437-440, 2006.

[26] P. D. Agnello, " Process Requirements for Continued Scaling of CMOS - The Need and Prospects for Atomic-level Manipulation," *IBM Journal of Research and Development*, vol. 46, no. 2/3, pp. 317-338, 2002.

[27] International Technology Roadmap for Semiconductors, 2005 Edition, (http://www.public.itrs.net/).

[28] S. -D. Kim, C. -M. Park, and J. C. S. Woo, "Advanced Model and Analysis of Series Resistance for CMOS Scaling Into Nanometer Regime—Part I: Theoretical Derivation", *IEEE Transactions on Electron Devices*, vol. 49, no. 3, pp. 457-466, 2002.

[29] S. -D. Kim, C. -M. Park, and J. C. S. Woo, "Advanced Model and Analysis of Series Resistance for CMOS Scaling Into Nanometer Regime—Part II: Quantitative Analysis", *IEEE Transactions on Electron Devices*, vol. 49, no. 3, pp. 467-472, 2002.

[30] S. D. Kim, S. Narasimha, and K. Rim, "An Integrated Methodology for Accurate Extraction of S/D Series Resistance Components in Nanoscale MOSFETs," *IEEE IEDM Tech. Digest*, pp. 149-152, 2005.

[31] H. Kam, L. Chang and T. -J. King, "Impact of Source-Drain Doping Profiles and Contact Schemes on FinFET Performance in the Nanoscale Regime," *IEEE Silicon Nanoelectronics Workshop Proc.*, pp. 9-10, 2004.

[32] A. Dixit, A. Kottantharayil, N. Collaert, M. Goodwin, M. Jurczak, and K. De Meyer, "Analysis of the Parasitic S/D Resistance in Multiple-Gate FETs", *IEEE Transactions on Electron Devices*, vol. 52, no. 6, pp. 1132-1140, 2005.

Chapter 2

Formation of Ultra-Shallow Junctions (USJs)

2.1 Introduction

Current trends of complementary metal oxide semiconductor (CMOS) technology scaling are diversified in many ways to meet the historical pace of intrinsic transistor performance improvement by about 17% per year. At each node there always existed severe skepticism of the ability to realize future generations by simply scaling the geometry of the devices. However, the technologists almost always managed to scale the devices with minimal change to the conventional CMOS process flow.

The biggest scaling challenge is to keep the short-channel effects (SCE) under control which manifests into degraded device performance. SCE is the decrease of MOSFET threshold voltage (V_{TH}) as physical gate length (L_g) is reduced. Figure 2.1 highlights the short-channel V_{TH} roll-off against L_g. The decrease in V_{TH} with decreasing

L_g can be explained by significant capacitive coupling of other MOSFET terminals (source and drain) with respect to control terminal (gate) in short-channel devices. SCEs are more pronounced when high voltage bias is applied to the drain terminal of MOSFET. High drain bias reduces the potential barrier at the source end of the channel for carriers to flow into the channel, resulting in further reduction of V_{TH}. This effect is also known as drain-induced barrier lowering (DIBL). Figure 2.2 shows the measured I_d-V_g characteristics of long-channel (L_g = 1.0μm) and short-channel (L_g = 0.175μm) devices at different drain biases. For a long-channel device, subthreshold current is independent of drain bias. However, for a short-channel device, V_{TH} is reduced for high drain bias condition, resulting in higher off-state leakage.

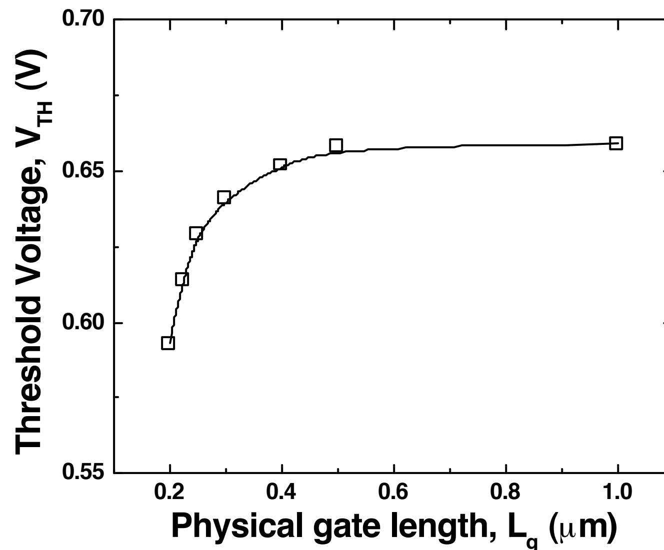


Figure 2.1 Linear V_{TH} roll-off characteristics for nMOSFETs with varying physical gate length. V_{TH} was extracted by peak G_m method using I_d-V_g measured at V_d = 50mV (details of MOSFET fabrication are discussed in section 3.4). EOT=1.15 nm.

When V_{TH} drops too much, off-state leakage becomes too large for that channel length, which is unacceptable. Thus, careful device design is needed so that V_{TH} roll-off and DIBL do not prevent the use of minimum physical gate length MOSFET. Using

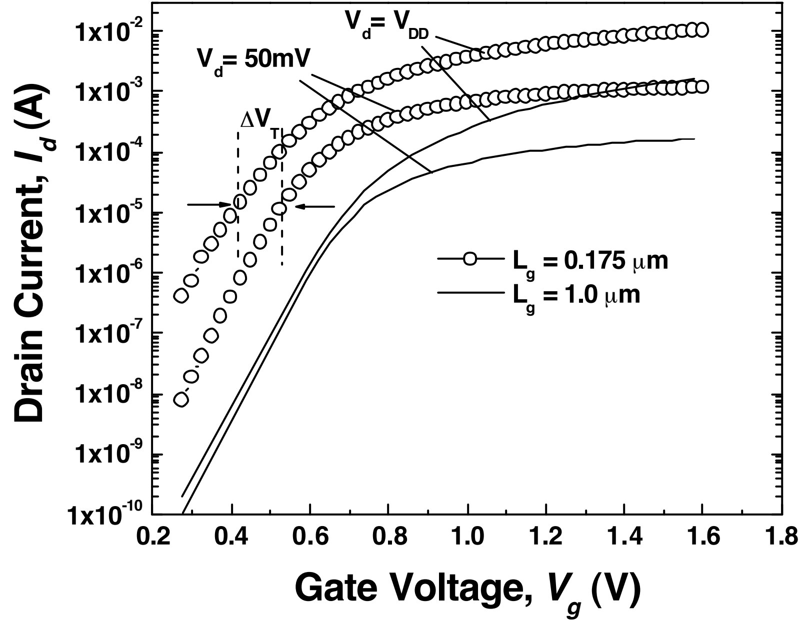


Figure 2.2 Subthreshold characteristics of a long-channel (L_g = 1.0μm) and short-channel (L_g = 0.175μm) nMOSFET. I_d-V_g was measured at V_d = 50mV and V_d = 1.2 V. DIBL for short-channel MOSFET is 100mV/V whereas DIBL for long-channel MOSFET is very small (10mV/V). EOT=1.15 nm.

Voltage-Doping Transformation, the shrinkage in MOSFET geometry can be correlated to V_{TH} roll-off and DIBL according to following equation [1] ?

$$SCE = \frac{\varepsilon_{Si}}{\varepsilon_{ox}}\left(1+\frac{X_j^2}{L_{el}^2}\right)\frac{T_{ox}}{L_{el}}\frac{T_{dep}}{L_{el}}V_{bi}$$
$$DIBL = \frac{\varepsilon_{Si}}{\varepsilon_{ox}}\left(1+\frac{X_j^2}{L_{el}^2}\right)\frac{T_{ox}}{L_{el}}\frac{T_{dep}}{L_{el}}V_d \qquad (2.1)$$

where L_{el} is the electrical channel length, X_j is the source/drain extension depth, T_{ox} is the gate dielectric thickness, T_{dep} is the depletion depth, ε_{Si} is the dielectric constant of silicon, ε_{ox} is the dielectric constant of gate dielectric, V_{bi} is the built-in potential, and V_d is the drain bias. Figure 2.3 shows the schematic cross-sectional of a MOSFET with abovementioned device dimensions.

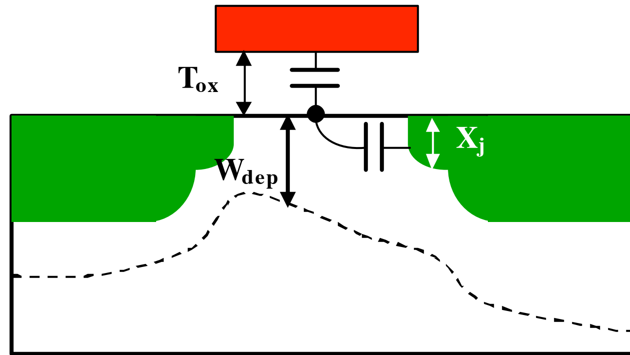


Figure 2.3 Cross-section of a MOSFET.

Eqn. (2.1) indicates that to avoid excessive SCE, vertical dimensions (T_{ox}, T_{dep}, X_j) must be reduced to support the gate length shrink. Reducing T_{ox} increases gate-to-channel capacitive coupling. Historically in bulk Si MOSFETs, thinning down the gate dielectric has been the key enabler to minimize SCE (and to improve drive current), as T_{ox} has been scaled by ~0.75× in each technology generation every three years as shown in Figure 2.4 [2]. However, recently the thickness of gate oxide has scaled more slowly than the historical pace and an equivalent oxide thickness (EOT) of ~1nm has been used for the past 2-3 generations, signaling the advent of new era where further scaling must be realized mainly by new materials. Nearly a decade of research and development has been conducted, by various groups, on high-permittivity (high-κ) gate dielectrics to potentially replace the conventional gate dielectric, SiO_xN_y, for future generations of CMOS. There has been excellent progress in recent years in demonstrating high-performance devices with hafnium-based high-κ dielectric and metal gate electrode [3]. Introduction of metal gates mitigates problems associated with conventional poly-silicon gate electrodes. While the introduction of high-κ/metal gates have extended effective gate oxide thickness scaling for better gate control, it will be difficult to scale EOT well below 1 nm.

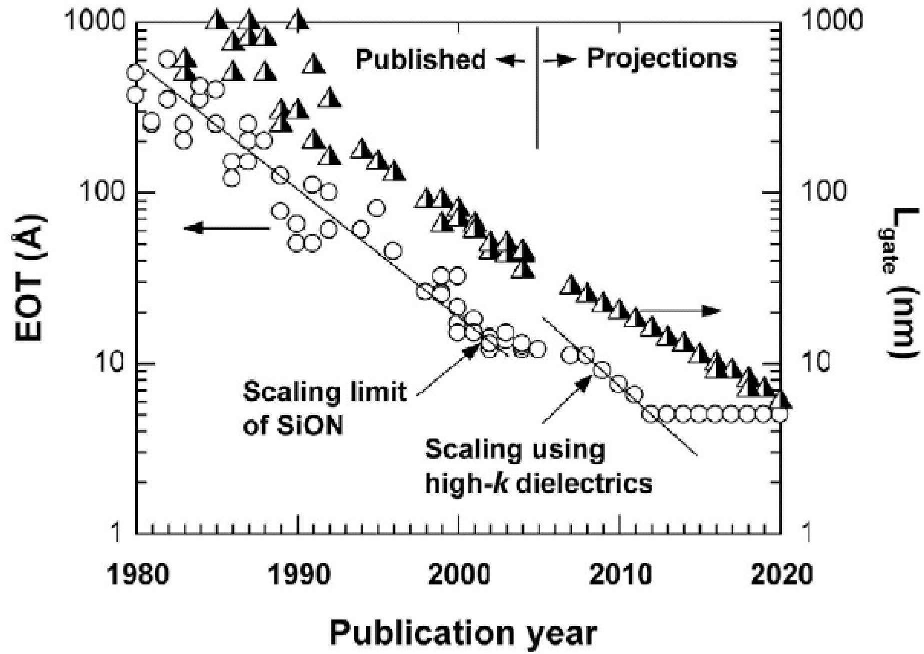


Figure 2.4 Published and projected scaling trend of equivalent gate oxide thickness (EOT) [2].

Reducing T_{dep} helps to eliminate leakage paths far from the dielectric-channel interface. T_{dep} can be reduced by increasing channel doping concentration but it also results in decrease carrier mobility due to enhanced impurity scattering degrading on-state performance. For current and near-future technology nodes, channel doping is ~4-6×10^{18} cm^{-3}, which is already too high [4].

T_{ox} and T_{dep} are clearly reaching fundamental physical limitations, which necessitates more aggressive junction depth scaling in order to control short channel effects. Figure 2.5 shows the X_j requirements for a given gate length for past and future technology generations to keep pace with historical improvement in high performance logic devices [4, 5]. In the 2005 edition, the *International Technology Roadmap for Semiconductors* (ITRS) model for X_j was ~$0.36\times L_g$. However, introduction of high-

κ/metal gate and offset spacers has relaxed this requirement and ITRS 2007 edition modifies it to be ~$0.50 \times L_g$. Nevertheless, future CMOS generations 32nm and beyond will have source/drain extensions few nm deep (~10nm), yet will require sheet resistance of the order of ~1000Ω/ .

The aforementioned junction requirements may appear to be too aggressive for realistic implementation in nanoscale devices, yet they stress upon the need to form ultra-shallow junctions (USJs) to keep short channel effects and total parasitic resistance of the device within acceptable degree of tolerance. The goal of this study is to systematically investigate the possibility of achieving shallower junctions with lower sheet resistance by conventional and/or advanced processing methods.

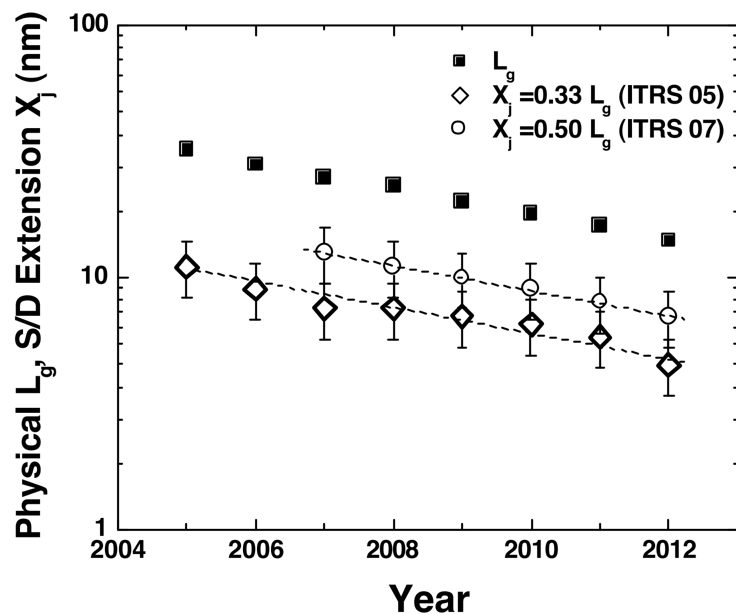


Figure 2.5 Scaling trend of physical gate length and source/drain extension as reported in ITRS.

2.2 State-of-the-art Junction Technology

In recent years, the formation of shallow junctions has seen a rise in advanced processes, at the cost of added process complexity, that accompany the conventional

method of dopant incorporation (by ion implantation) and dopant activation with implant damage annihilation (by rapid-thermal processing). Particularly, requirements for $p+/n$ shallow junctions are difficult to meet as compared to $n+/p$ junctions because of an extended tail of the boron (p-type impurity) profile due to ion channeling. Pre-amorphization implants (PAIs) such as germanium or silicon are primarily used to minimize ion channeling. If silicon is amorphized during implantation, extended defects are formed upon annealing. These defects are commonly known as end-of-range (EOR) defects and are located just below the interface between the amorphized and crystalline substrate. EOR defects are typically {311} rod-like defects or dislocation loops, formed by excess silicon interstitials (Si^i). If the EOR damage layer lies within the depletion width, junction leakage is increased. Hence, the amorphized silicon depth should be carefully controlled to avoid the presence of EOR damage within the depletion width (which depends upon source/drain extension and halo doping). Since boron diffuses in silicon by interstitial mechanism, excess Si^i concentration leads to an enhanced diffusion of boron commonly known as transient enhanced diffusion (TED) [6]. While the dopant profile may be shallow after implantation, upon annealing the profile can diffuse several nanometers into the bulk substrate resulting in undesired deep junctions. Hence, the control of junction depth for a given technology node is limited by TED.

It has been demonstrated that co-implantation of carbon and fluorine can be a very effective method to reduce TED of boron atoms [7, 8]. These co-implanted species form complexes with Si^i, which reduces $[Si^i]$ to near-equilibrium values, reducing B-Si^i interaction for diffusion. Best results are obtained when projected range (R_p) of co-implanted species is close to the amorphous (a-Si)/crystalline (c-Si) interface generated by

pre-amorphization as shown in Figure 2.6, preventing the arrival of Si^i in the boron-doped region. This results in a twofold advantage – shallower junctions due to lesser boron TED and stabilization against deactivation caused by interstitial-driven boron clustering [9, 10]. Thus, there is a strong interdependence between the depth of amorphized silicon region and co-implanted species profile. Therefore both – PAI which determines the depth of amorphized silicon and co-implanted species' implanted profile need careful investigation to achieve optimum junction depth. In addition to carbon and fluorine, there are also reports of the use of nitrogen as a co-implant species to control boron TED [11, 12].

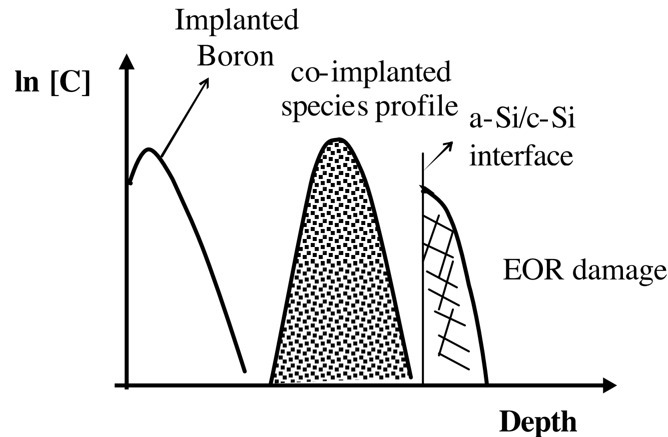


Figure 2.6 To mitigate boron TED, co-implanted species energy is chosen in a way so that the projected range of co-implantation lies in-between the implanted boron profile and a-Si/c-Si interface. EOR damage layer is located beneath this interface.

To further reduce the impact of TED, it is necessary to anneal at the highest temperature possible while limiting the total thermal budget for dopant diffusion. Consequently, spike annealing with soak time less than 1 second has been developed. Fast-ramp spike annealing along with low-energy dopant implants with co-implantation (for boron) is currently in practice to satisfy the shallow junction requirements for nodes down to 65 and 45 nm.

2.3 Alternate Doping and Annealing Techniques

The demand for low-energy implantation will be increasingly difficult to meet for 32 nm nodes and below because throughput is limited by low beam current from the ion source to wafers. To address the issue of decreased productivity at low energies, decel mode is used. The ions are extracted at higher energies to draw high beam current and then decelerated before they reach the wafer targets. However, some of the dopant ions are neutralized by charge exchange with the background gas prior to deceleration. These neutrals will not be decelerated and therefore reach the wafer at higher than desired energies. This effect is known as energy contamination and leads to a deeper than desired dopant depth profile. In decel mode operation, beam spreading becomes a problem because of increased coulomb repulsion due to decreased ion velocities for low energy implants. These two effects are shown in Figure 2.7 [13].

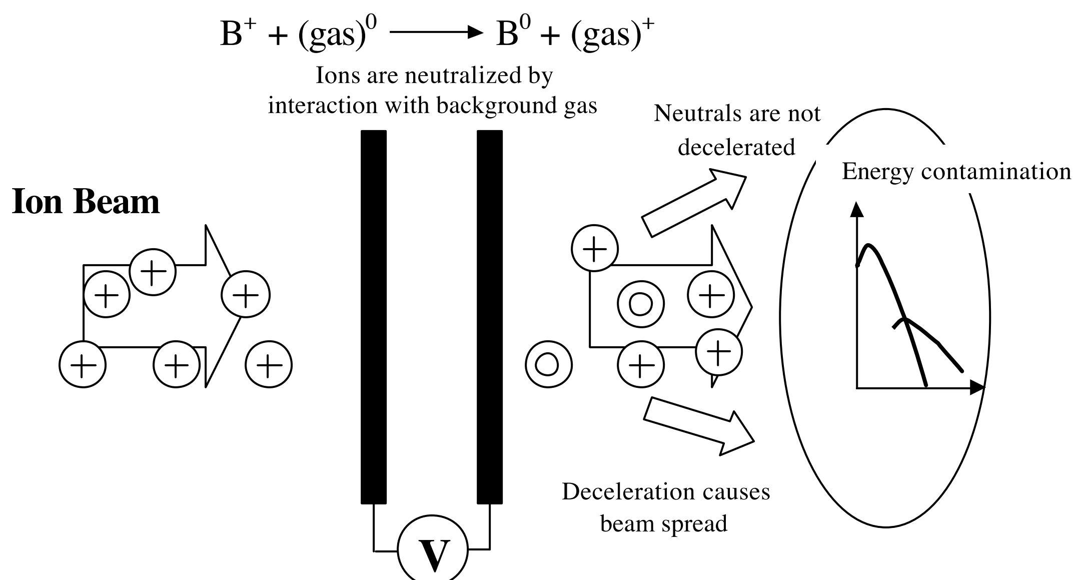


Figure 2.7 Mechanism of energy contamination in decel mode ion implanter. Energy contamination causes implanted profile to be deeper than desired.

Various doping techniques such as Plasma doping, Infusion doping, and Molecular implants have been proposed to counter the limitations of beam-line implantation. In the PLAsma Doping (PLAD) technique, the wafer is placed in a vacuum chamber filled with process gas (BF_3, AsH_3, PH_3). Application of negative voltage to the wafer ionizes the gas and attracts these ions toward the wafer, resulting in ion implantation. In PLAD, all ionized species present in the plasma will be implanted into the wafer, causing the risk of contamination and etching. To minimize these effects, a series of voltage pulses is applied to the wafer such that plasma is present only during the implant time. This modified process is called Pulsed PLAsma Doping (P^2LAD) [14]. Within-wafer and wafer-to-wafer uniformity are important issues for PLAD because implanted ions have a range of energies rather than a specific energy[15].

Infusion doping utilizes gas cluster ion beam (GCIB) for shallow doping applications [16, 17]. GCIB processing uses clusters (>5000 atoms) consisting of Ar or He and B containing molecules such as B_2H_6 and BF_3. After adiabatic expansion that causes cluster formation, clusters are ionized and accelerated to the substrate. Upon impact, the cluster energy is transferred to the substrate surface causing rapid heating and dissociation of molecules into its atoms. The gas atoms (Ar, He, H) leave the surface and soluble species intermix or are infused in the substrate. Although the clusters have higher total energy, the energy/atom is very low (~10eV) and hence intermixing occurs in very shallow depths. Infusion doping is mass independent because the intermixing depth is determined by the collective energy of the cluster. Extreme abruptness and lack of channeling or energy contamination are the main features of infusion doping.

The aforementioned doping techniques are promising for shallow junction applications but require in-depth research on various aspects of device performance before they can be implemented in production. Furthermore, they have to be cost-effective and should obtain same or better process control as achieved by conventional method of doping – beam-line implantation. Thus alternative methods are being explored to extend conventional beam-line implantation to several more generations. Molecular implants such as $B_{10}H_{14}$ and $B_{18}H_{22}$ are well suited for direct replacement of monoatomic ion species for low-energy implantation providing improved throughput while retaining the precision of ion dose, repeatability, uniformity and angle control of ion implantation [18, 19].

Rapid thermal processing is also facing contrasting challenges of dopant movement during annealing and limited electrical activation because of solid solubility limits. A higher anneal temperature is desirable but with current RTA tools, the annealing time at these elevated temperatures are too long to cause significant dopant diffusion, resulting in deeper junctions. Figure 2.8 indicates that in order to limit diffusion to <2nm, milli-second anneal will be needed for $T>1250^{o}C$.

A spike anneal with a very high ramp rate has been used to meet 45 nm CMOS technology requirements for shallow junctions. But in order to meet the requirements for sub-45 nm technologies, spike annealing will not be adequate; especially, boron TED will still pose great challenges for ultra-shallow junction formation. Flash annealing and laser annealing are two prime candidates for possible replacement of spike annealing and are often regarded as "diffusion-less" annealing techniques. Flash annealing uses an array of flash lamps ignited by high voltage pulses to achieve annealing times in the range of

milli-seconds [20, 21]. Laser annealing is a metastable process lasting few microseconds to nanosceconds in which dopants can be frozen in the lattice sites well above the solid solubility limit [22], but equipment maturity is the main challenge for its adoption. Both annealing techniques are capable of heating a silicon substrate to very high temperature. Figure 2.9 exhibits the differences in annealing temperature and time for conventional anneals (furnace, RTP, spike) and advanced anneals (flash and laser). With advanced annealing techniques, it is possible to achieve higher temperature (>1200°C) for increased dopant activation while keeping the anneal time short for restricted dopant diffusion. Another primary difference is that advanced annealing techniques heat up only the device side of the wafer, unlike conventional annealing techniques for in which the bulk of the wafer is subjected to thermal treatment.

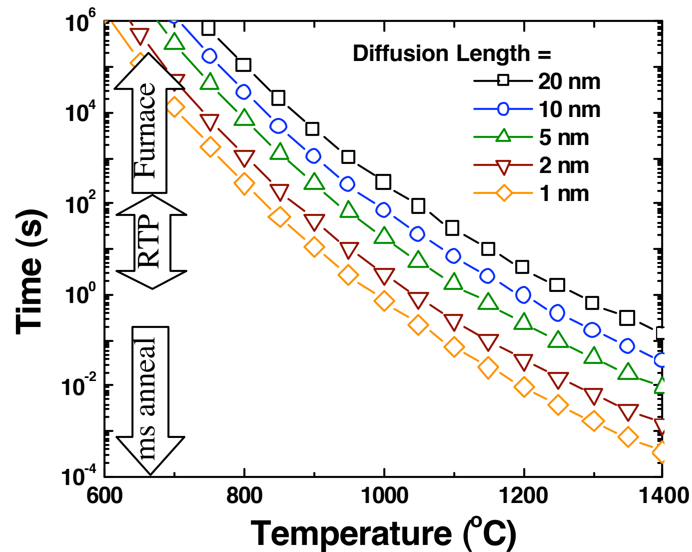


Figure 2.8 Thermal budget criteria for different B diffusion lengths 2√Dt). High-temperature milli-second anneals can provide for higher dopant activation without significant diffusion.

Solid phase epitaxial regrowth is another potential candidate to achieve shallow junctions. The main advantage is above-equilibrium dopant activation, good control over

junction depth and lower temperature requirement (~650°C) [23]. The main drawback is relative high density of residual defects after annealing which leads to very high junction leakage.

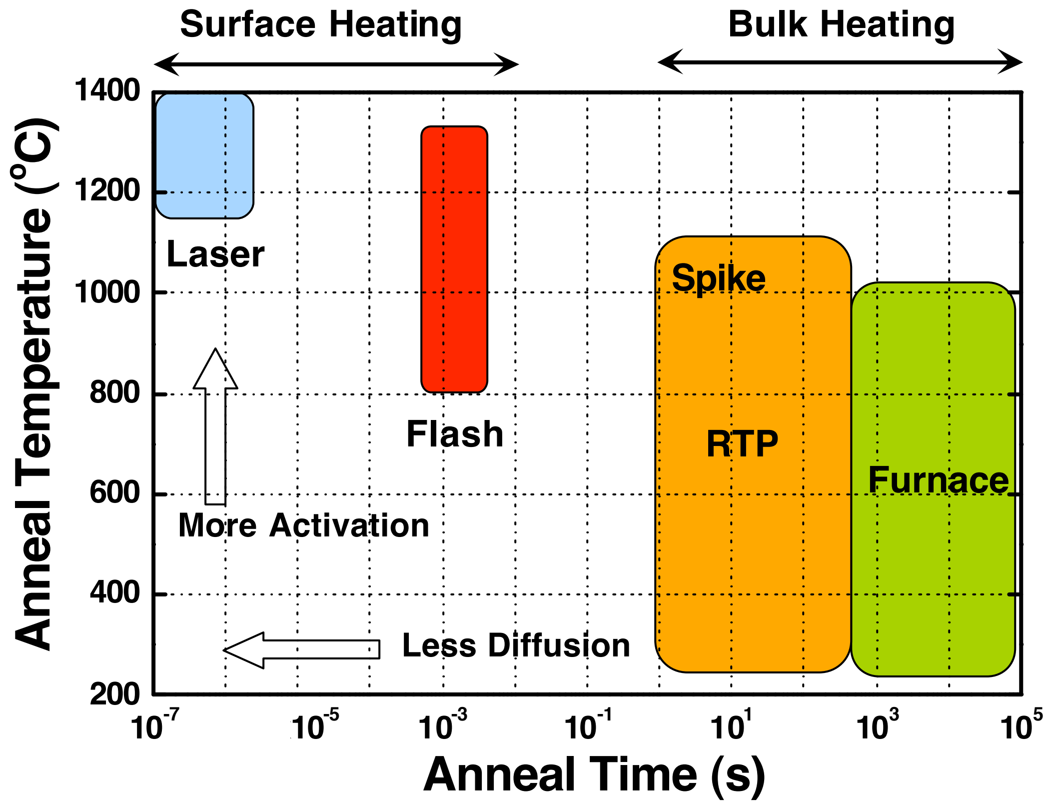


Figure 2.9 Temperature-time ranges of various conventional and advanced annealing techniques.

2.4 Junction Metrology

In this work, Secondary Ion Mass Spectroscopy (SIMS) was used to measure the junction depth (X_j). The analyses were performed on the Physical Electronics ADEPT-1010. B was monitored as a positive ion under 500-650eV O_2^+ bombardment incident at 45°. The analysis chamber was backfilled with a high partial pressure of O_2 to reduce ion intensity fluctuations at the beginning of the profile. As was monitored as the molecular secondary ion As+Si under 500eV Cs^+ bombardment at 60°. Secondary ions were

collected from the center 10% of a 400 μm x 400 μm rastered area. 1-cm^2 pieces were taken from the center of each wafer. Stylus profilometry was used to determine the depth of the craters and calibrate the depth scale. B and As concentrations were calculated using relative sensitivity factors (RSFs) determined from standard samples.

Accurate sheet resistance (R_s) measurement is critical for assessing USJ formation processes. The conventional method, four-point probe (FPP), is used to measure sheet resistance. However, a mechanical load is used to push the probes against the sample, which causes probe penetration and leaves a footprint behind because of silicon fracture. This probe-tip penetration can be deeper than X_j for USJs as shown in Figure 2.10(a). Figures 2.10 (b) & (c) are atomic force microscopy (AFM) topographic images of samples probed using two FPP tools [24]. It is clear that both tools penetrate significantly into the silicon surface; the largest penetration depth is 30-130 nm.

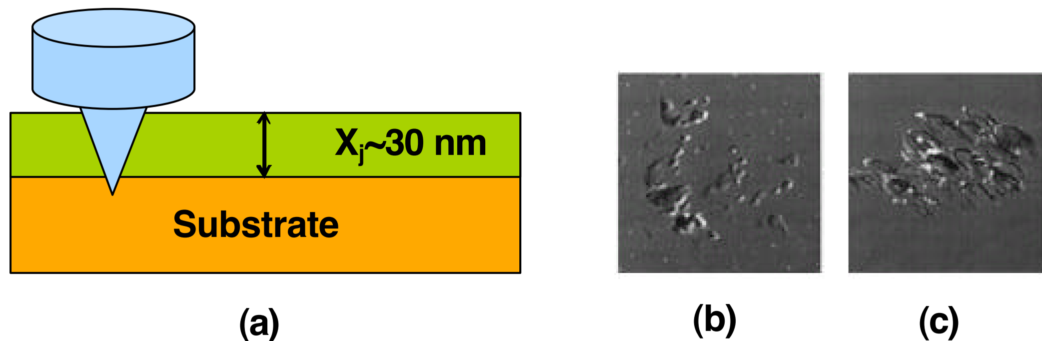


Figure 2.10 (a) Probe-tip penetration for junctions shallower than 30nm in FPP method (Source: Dr. M. Current). **(b), (c)** AFM images of the probe imprints of FPP SSM 240 (image size 5 μm) and FPP Veeco 5000 (image size 50 μm), respectively. The z ranges are 100 and 500 nm [24].

Sheet resistance was also measured by a non-contact method for improved accuracy. The measurement is based on the monitoring of junction photo-voltage at a transparent electrode at the center of the probe (V_{in}) and second electrode (V_{out}) some

small distance away as shown in Figure 2.11 [25, 26]. Free carriers are created in the depletion region by illuminated light. These carriers, then, get separated at the junction boundary. The charge in the surface junction induces a charge at the central electrode. This is the junction photo-voltage (JPV) signal. Carriers created in the junction spread out from the illuminated area and a JPV signal is detected by the outer electrode. The ratio and phase of the JPV signals, V_{in}/V_{out} is used to determine R_s.

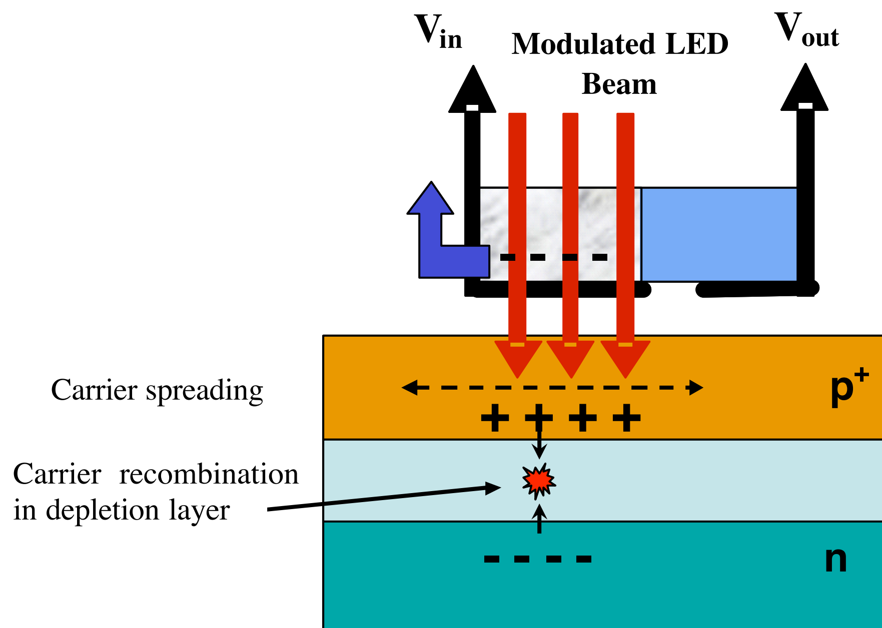


Figure 2.11 Non-contact method of measuring sheet resistance and leakage of a *p/n* junction [25].

The non-contact method can also be used to measure junction leakage. Carriers which drift into the depletion layer can recombine at defect sites. This recombination reduces the charge in the surface junction and JPV signal. Analysis of the JPV signal at different modulation frequencies allows for determination of junction recombination current density.

Variability for R_s measurements using the non-contact method is 0.05% as shown in Figure 2.12 (a) [27]. The other advantage of the non-contact method is that whole wafer mapping can be done very quickly (~1,000 pt in <2.5 min). A R_s map for one of the studied wafers is shown in Figure 2.12 (b).

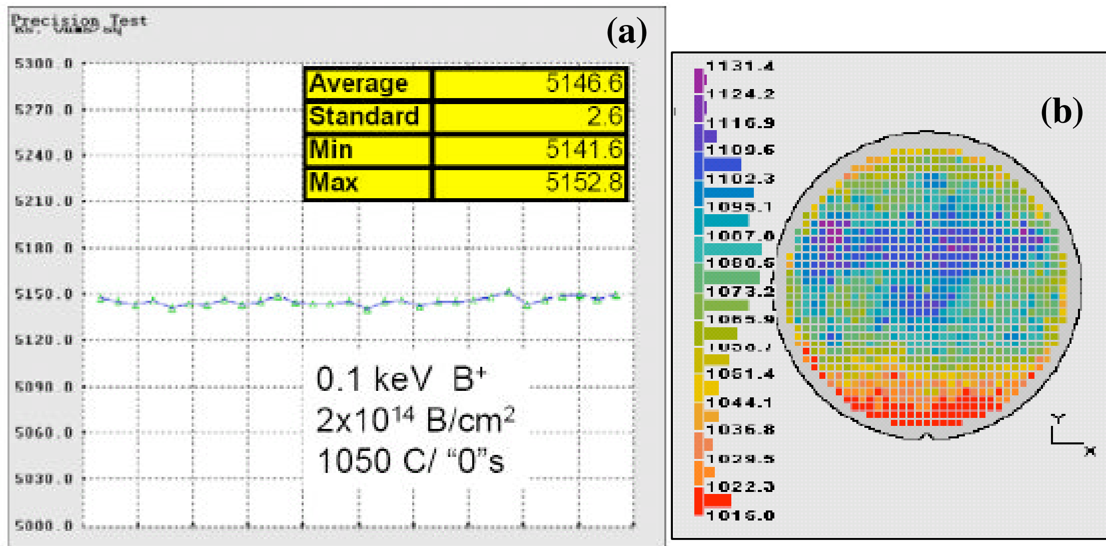


Figure 2.12 A 30-point repeatability test of sheet resistance measurement 27]. **(b)** Whole wafer sheet resistance mapping.

Measurements obtained with the conventional contact method *vs*. non-contact method are compared in Figure 2.13. The conventional method causes leakage error in R_s measurements due to probe-tip penetration to a depth lower than X_j, yielding erroneously low values of measured R_s. The error increases exponentially as X_j decreases, exceeding 1000 Ω/ for X_j < 15nm. Therefore, the non-contact measurement is employed in this work for accurate USJ characterization.

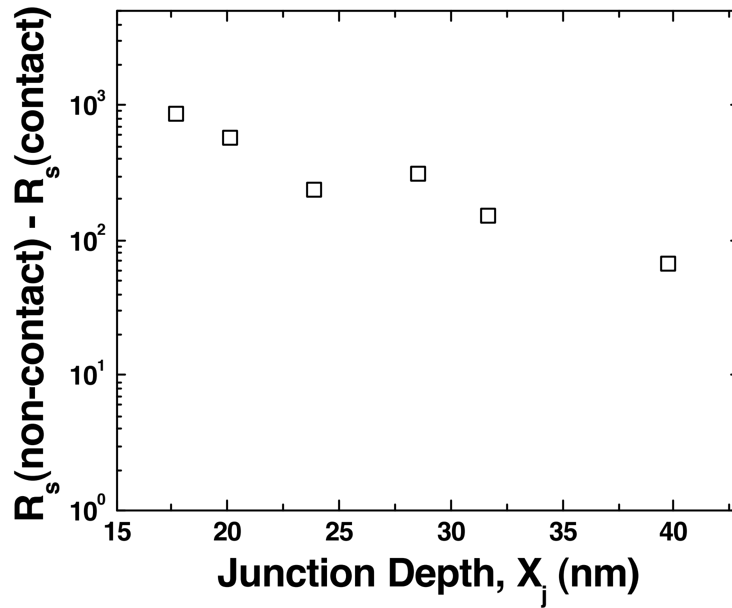


Figure 2.13 Difference between measured sheet resistance values for the non-contact method and the conventional contact method.

2.5 Experimental Results with Spike Anneal

Blanket USJ formation studies were performed using lightly doped (~10^{15}cm^{-3}) *n*- or *p*-type Si (100) wafer substrates. 2nm-thick thermal oxide was grown before ion implantation, followed by either shallow B (0.2keV, 5×10^{14} cm^{-2}, 7°) or BF_2 (0.5-1.0keV, 5×10^{14} cm^{-2}, 7°) implant for *p+/n* junctions and shallow As (0.5-1.0keV, 10^{15} cm^{-2}, 7°)

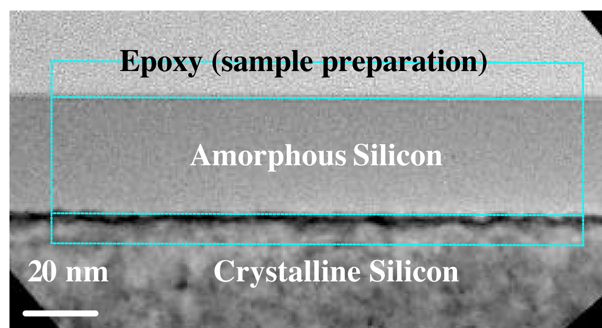


Figure 2.14 Bright field transmission electron microscopy (BFTEM) image of cross-section of amorphized silicon sample. The amorphous Si layer is 36 nm thick.

implant for $n+/p$ junctions. A pre-amorphization Ge implant (PAI) (20keV, 10^{15} cm^{-2}) was done prior to boron implantation. The amorphized silicon depth was 36 nm as shown in the cross-sectional transmission electron microscopy image (Figure 2.14). SIMS analysis results of the as-implanted boron profiles (BF_2 = 0.8keV) with and without PAI Ge are shown in Figure 2.15. PAI helps to eliminate channeling, which results in an extended tail below ~10^{19} cm^{-3} in the absence of PAI.

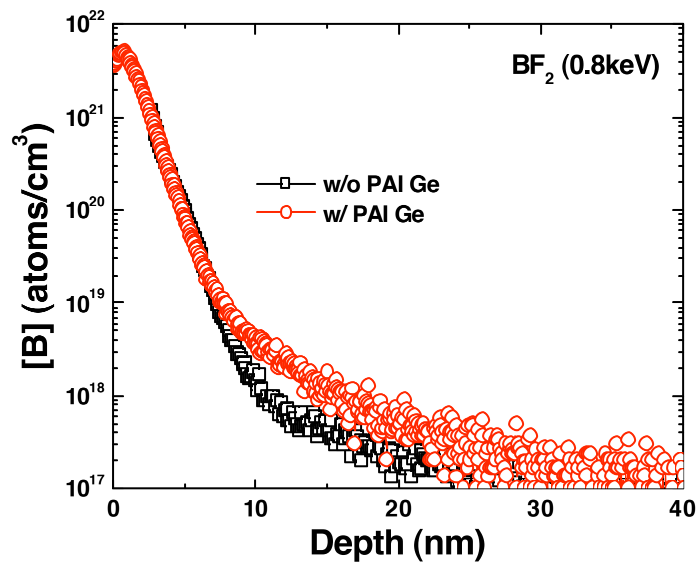


Figure 2.15 SIMS profiles with and without pre-amorphization implant (PAI). An extended tail is seen in the absence of PAI.

Nitrogen (5 or 9keV, 10^{15} cm^{-2}) and fluorine (6 or 11keV, 10^{15} cm^{-2}) were co-implanted with boron on some samples. The implant energies of these co-implanted species were chosen such that their R_p is in-between the implanted boron and the amorphous/crystalline interface. A spike anneal (1070°C) was used to activate the implanted dopants.

SIMS analysis results for low energy boron- and arsenic-implanted samples before and after spike anneal are shown in Figure 2.16 and Figure 2.17. As-implanted

junction depths for these profiles are less than 10nm. (The junction depth X_j is taken to be the depth at which the dopant concentration falls to 5×10^{18} cm^{-3}.) Spike annealing causes excessive diffusion so that the junction depth for both boron and arsenic after annealing is >25nm. Thus, spike annealing results in junctions that are too deep for sub-45nm nodes.

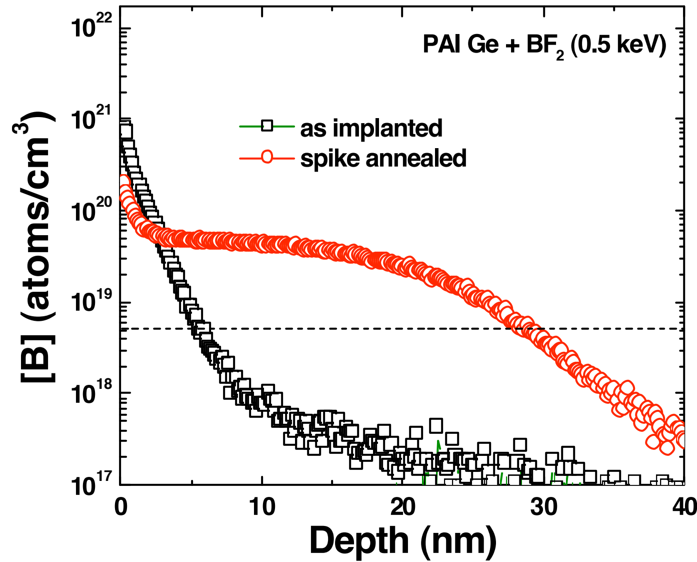


Figure 2.16 B SIMS profiles of as-implanted and spike-annealed for BF_2 implant (0.5keV, 5×10^{14} cm^{-2}). Significant dopant diffusion is seen for the spike anneal. Junction depth is taken at [B] = 5×10^{18} cm^{-3}.

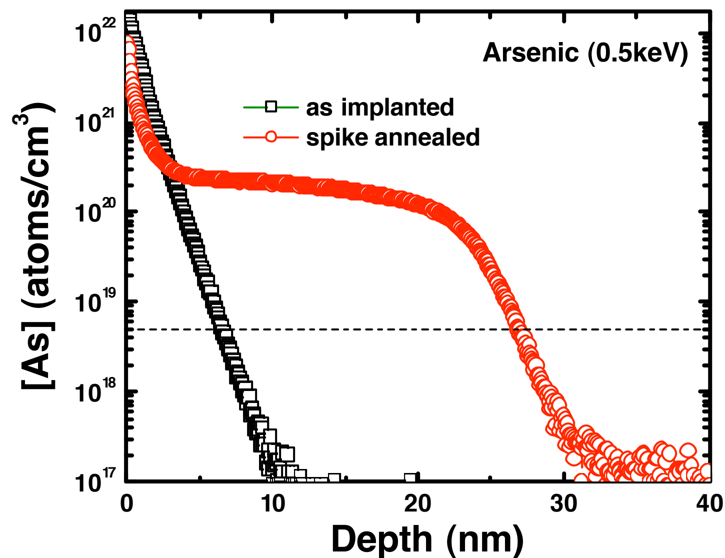


Figure 2.17 As SIMS profiles of as-implanted and spike-annealed As implant (0.5keV, $10^{15}cm^{-2}$). Spike annealing results in much deeper junctions (X_j>25nm).

Co-implanted species interact with excess Si interstitials generated during implantation and thus mitigate boron TED. Figures 2.18 & Figure 2.19 highlight the impact of F and N co-implantation, respectively, for different implanted boron energies.

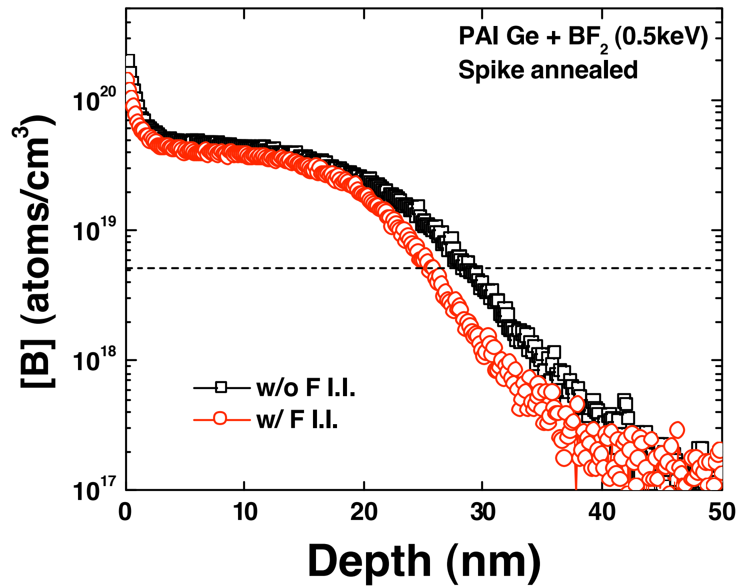


Figure 2.18 SIMS profiles of spike annealed BF_2 implant (0.8keV, 5×10^{14} cm^{-2}) without and with F co-implantation.

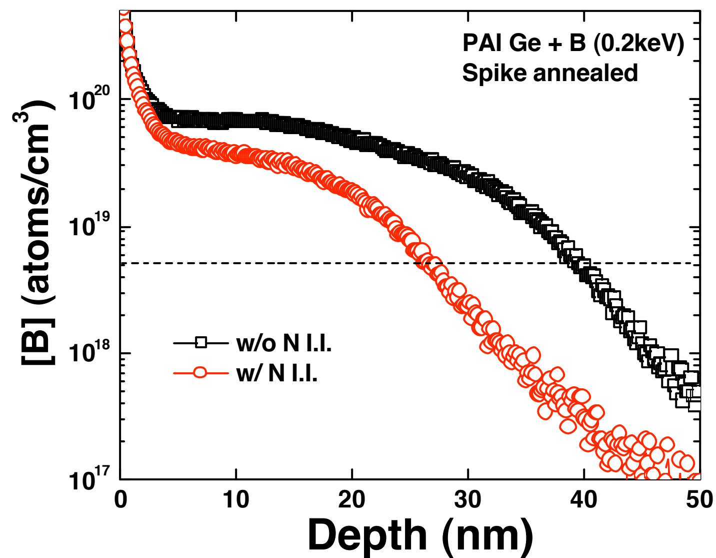


Figure 2.19 N co-implantation helps to reduce X_j, as seen in the SIMS profiles of spike annealed B (0.2keV, 5×10^{14} cm^{-2}).

The total change observed in junction depth (ΔX_j) is 3-10nm for all studied co-implantation conditions. Therefore, co-implantation helps to reduce X_j, but this reduction is not enough to meet ITRS specifications for highly scaled CMOS technologies.

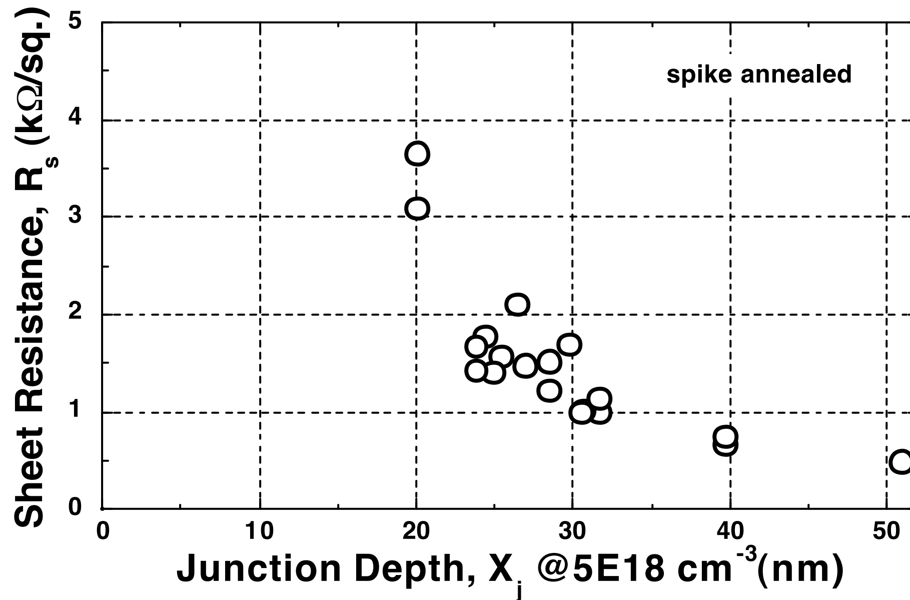


Figure 2.20 Sheet resistance (R_s) determined by non-contact method *vs.* junction depth (X_j) for *p+/n* junctions formed by spike annealing (1070°C).

Figure 2.20 summarizes the experimental results obtained with spike annealing with and without co-implantation for *p+/n* junctions. These results indicate that it may be possible to achieve junction depths of ~20nm, but it comes at the expense of higher sheet resistance. Furthermore, it will be increasingly difficult to meet ultra-shallow junction requirements below 20nm with spike annealing while maintaining low sheet resistance.

2.6 Experimental Results with Laser Anneal

The substrate used in this study was *n*-type (100) silicon. The samples were first pre-amorphized using Ge implantation, which resulted in 36 nm of a-Si. Subsequently, BF_2 implant was carried out at 0.5 keV with a dose of 5×10^{14} cm^{-2} and angle of 7°. A pulsed XeCl excimer laser anneal (ELA) with λ=308nm was used to activate dopants.

The ELA pulse was of very short duration, approximately 30ns. Three different laser fluences were chosen in this experiment: 450 mJ/cm^2, 500mJ/cm^2, and 550mJ/cm^2. SIMS results obtained are summarized in Figure 2.21.

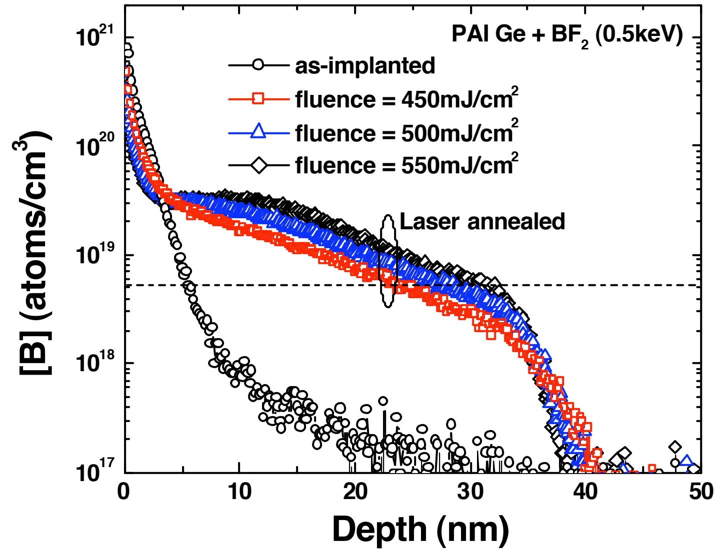


Figure 2.21 SIMS profiles of BF_2 (0.5keV, 5×10^{14} cm^{-2}) implant before and after laser anneal. Laser fluence of 450 mJ/cm^2, 500mJ/cm^2, or 550mJ/cm^2 was used to activate dopants.

ELA has near-zero thermal budget, but as shown in Figure 2.21, it results in deeper junctions (X_j > 20nm), and shows no significant benefit over spike annealing for this particular experiment. This anomaly can be attributed to the chosen laser fluence and depth of pre-amorphization, as discussed next.

The laser fluence needed to melt any material is defined as the melt-threshold of that material. Hence, laser annealing processes generally fall into two categories: melt or non-melt. Dopant diffusivity in molten silicon is about 8 orders of magnitude higher than that in solid-phase silicon. Therefore, dopants will diffuse very rapidly in molten silicon but slowly in solid-phase silicon, resulting in a step-like profile [28]. The melt-threshold also depends on the crystalline state since amorphized material has lower melting point as

compared to crystalline material. Hence, the melt-threshold of a-Si is much lower than c-Si.

In the experiments discussed above, the amorphized silicon depth is 36 nm and the chosen laser fluence is higher than the melt-threshold of a-Si but not enough to melt c-Si. Therefore, the laser pulse melts ~36nm of silicon and a step-like gradient is observed near 36 nm for all studied conditions. This observation indicates that the depth of pre-amorphizated silicon can set the junction depth and shallow junctions can be achieved by picking a lower PAI Ge energy as reported in [29].

2.7 Experimental Results with Flash Anneal

USJ studies were done using lightly doped (~10^{15} cm^{-3}) *n*- or *p*-type Si (100) wafer substrates with 2 nm-thick thermal oxide as capping layer. For *n+/p* junctions, As implantation at 0.5-1.0 keV with 10^{15}-2.5×10^{15} cm^{-2} dose and 7° tilt was performed. *p+/n* junctions were formed by BF_2 implantation (0.8-1.0keV, 5×10^{14} -2.5×10^{15} cm^{-2}, 7°). 36 nm of silicon was pre-amorphized prior to BF_2 implantation by Ge implant (20 keV, 10^{15} cm^{-2}).

Figure 2.22(a) highlights the features of the flash annealing process used in this work. In a flash annealing process, the bulk of the wafer is first heated up to an intermediate temperature (T_{int}), and then a millisecond flash is applied to heat the device side of the wafer to the peak temperature (T_{peak}). Once the flash lamps are turned off, the bulk acts as a thermal sink to cool the device side very rapidly, and finally radiative cooling to ambient temperature occurs. Large heating and cooling rates (~10^6 $^oC/s$) can be achieved using flash anneal. Figure 2.22(b) shows a typical heating cycle for a flash

anneal. High speed pyrometers are used to the monitor temperatures of the top and bottom surfaces of the wafer [30].

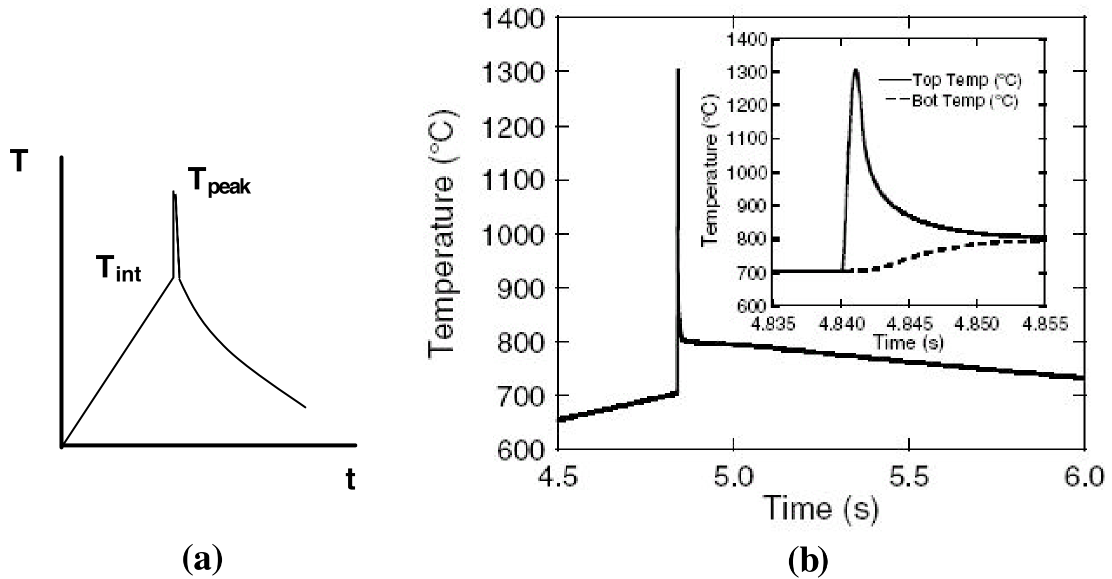


Figure 2.22 **(a)** Device (top) side temperature-*vs.*-time profile for flash annealing. The wafer temperature is first ramped up to an intermediate temperature (T_{int}), and then the device side of the wafer is heated to the peak temperature (T_{peak}). **(b)** Measured temperatures at the top and bottom surfaces of a wafer during flash annealing.

SIMS analyses of boron (BF_2=0.5 keV) and arsenic (As=0.5 keV) implanted samples before and after flash anneal are shown in Figure 2.23 and Figure 2.24, respectively. Spike annealed profiles are also plotted on the same scale for reference.

Negligible dopant movement is observed for flash annealed samples compared to spike annealed samples. Clearly, flash annealing reduces dopant diffusion as compared with spike annealing and is a promising approach for ultra-shallow junctions formation for sub-45nm nodes. These results are consistent with those of other studies [31, 32].

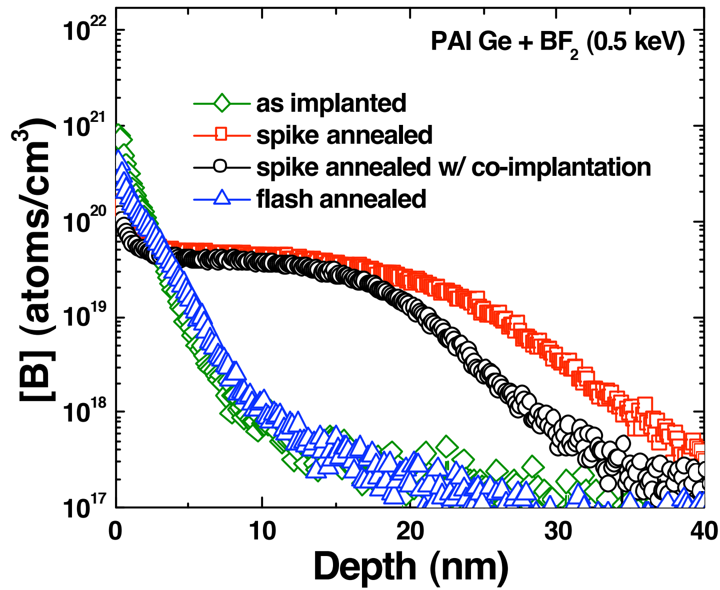


Figure 2.23 SIMS analyses of samples which received low-energy boron (0.5keV) implant. Negligible diffusion is observed for flash annealed samples compared to spike annealed samples (with or without co-implantation).

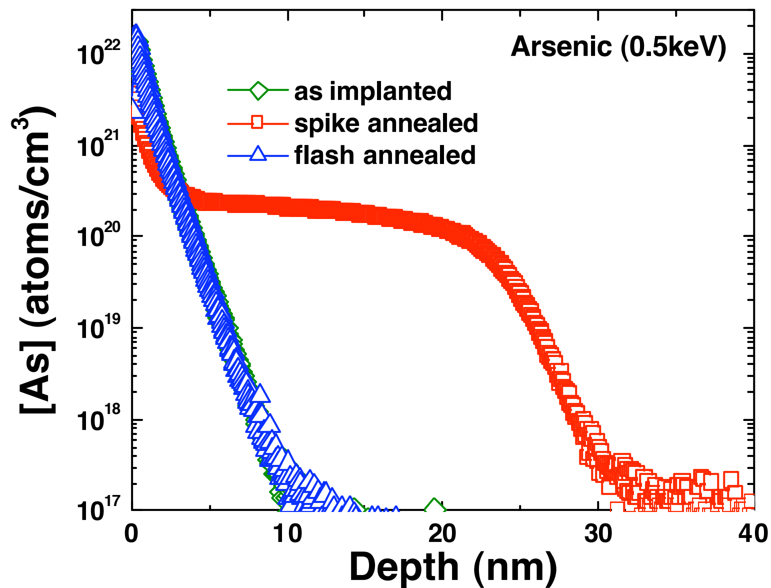


Figure 2.24 SIMS analysis for as-implanted, spike annealed and flash annealed samples implanted with As (0.5keV). Flash annealing results in restricted dopant movement as compared with spike annealing.

Various combinations of flash anneal parameters (T_{int}, T_{peak}, and n, where n is number of flash shots) were also included in the study to elucidate interdependencies.

The time between flash anneals was relatively long, for the samples which received multiple flash anneals, as illustrated in Figure 2.25. Figures 2.26-2.31 exhibit the effects of T_{int}, T_{peak}, and n on sheet resistance (R_s) and junction depth (X_j).

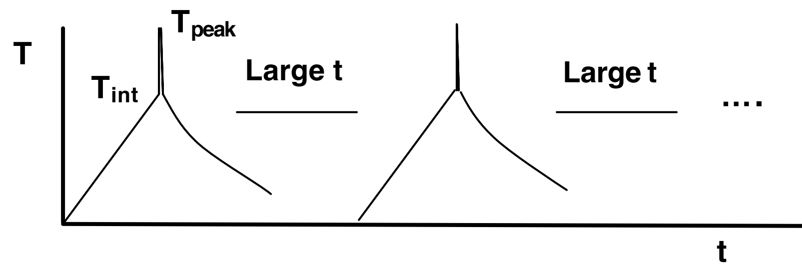


Figure 2.25 Temperature-*vs.*-time profile for multiple-shot flash annealing with relatively long time intervals in between.

Figures 2.26 & 2.27 show the impact of T_{int} for two different values of T_{peak}, for single-shot flash annealing. For both 1250°C and 1300°C peak temperature, T_{int} = 700°C results in shallower X_j and lower R_s. Thus, 700°C is preferred over 800°C for the intermediate temperature.

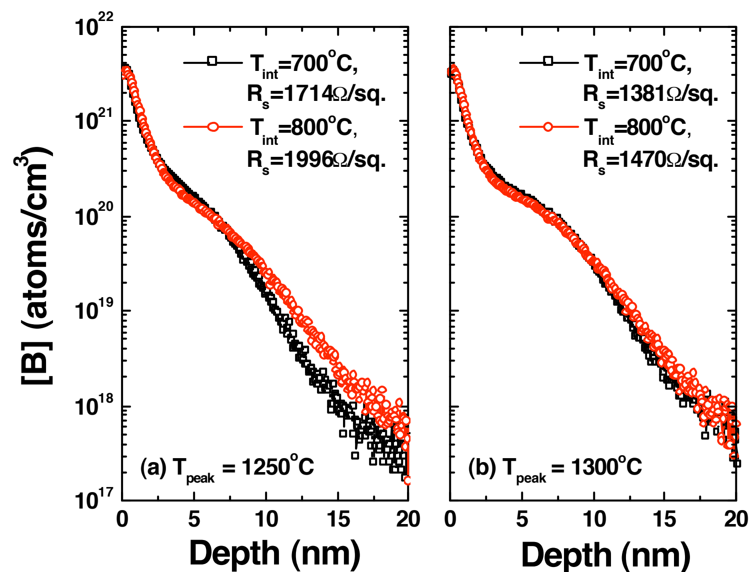


Figure 2.26 SIMS profiles of samples implanted with BF_2 (0.8keV, 10^{15} cm^{-2}) flash annealed at various T_{int} with **(a)** T_{peak} = 1250°C, (b) T_{peak} = 1300°C.

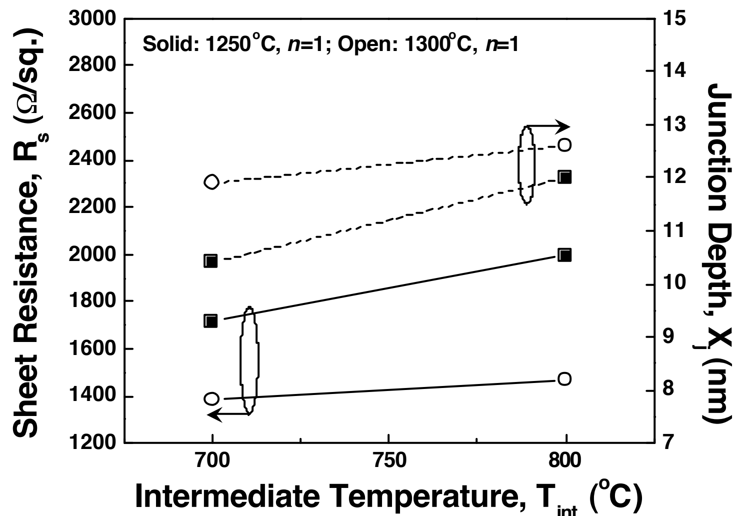


Figure 2.27 Measured R_s and X_j as a function of T_{int}. Solid symbols are for T_{peak} = 1250°C and n = 1. Open symbols are for T_{peak} = 1300°C and n = 1.

Figures 2.28 & 2.29 show the impact of T_{peak}. Significant reduction in R_{s} (26% for T_{int} = 700°C and 20% for T_{int} = 800°C) is observed when T_{peak} is increased from 1250°C to 1300°C. However, the reduction in R_{s} comes at the expense of a small increase in X_{j} (<1nm). The combination of T_{int} = 700°C and T_{peak} = 1300°C appears to be optimal for achieving shallow X_{j} with low R_{s}.

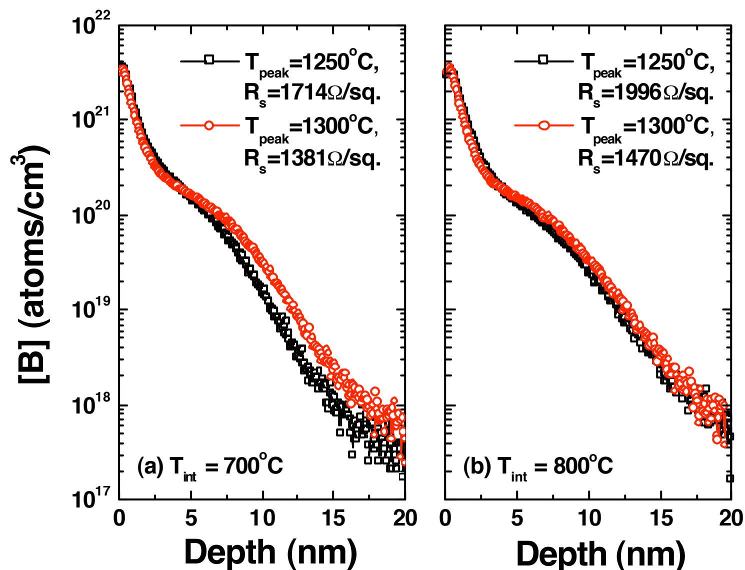


Figure 2.28 SIMS profiles of samples implanted with BF_2 (0.8keV, 10^{15} cm^{-2}) flash annealed with **(a)** T_{int} = 700°C, (b) T_{int} = 800°C, for two various T_{peak}.

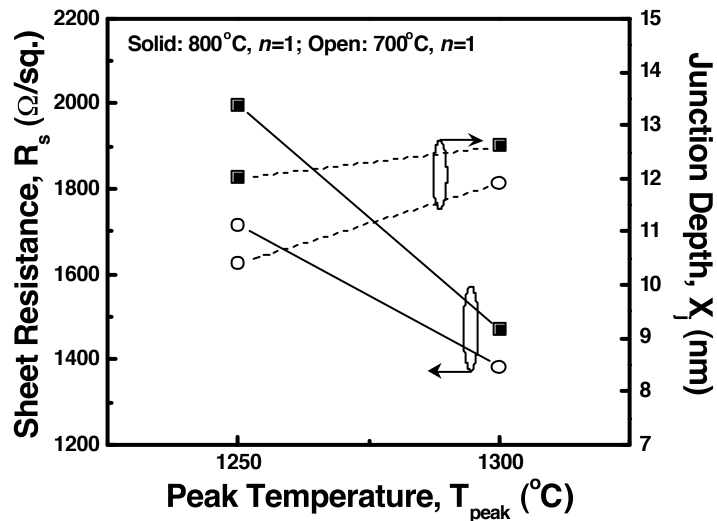


Figure 2.29 Measured sheet resistance and junction depth as a function of T_{peak}. Solid symbols are for T_{int} = 800°C and n = 1. Open symbols are for T_{int} = 700°C and n = 1.

Figures 2.30 & 2.31 demonstrate the impact of multiple shots of flash annealing. An additional 20% decrease in R_s is achieved as n is increased from 1 to 3. However, the junction depth is also increased, by ~1.5nm. Therefore, multiple shots of flash annealing can be used to further reduce R_s with a slight trade-off in X_j.

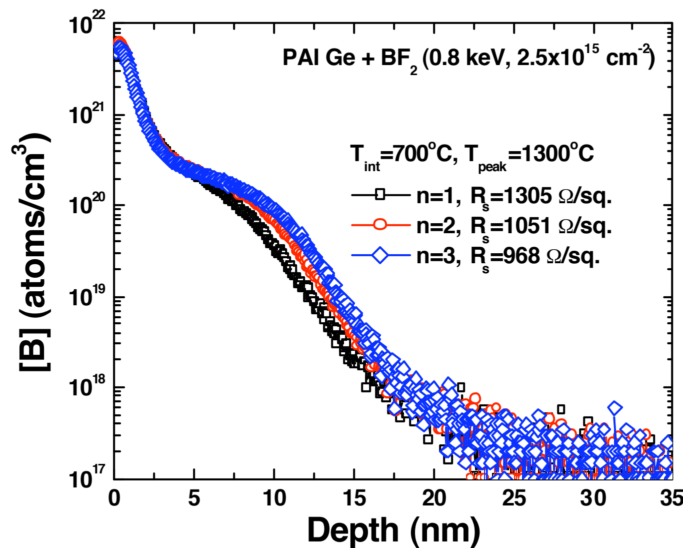


Figure 2.30 SIMS profiles of samples implanted with BF_2 (0.8keV, 2.5×10^{15} cm^{-2}), for various number of shots of flash anneal T_{int} = 700°C and T_{peak} = 1300°C. n is varied from 1 to 3.

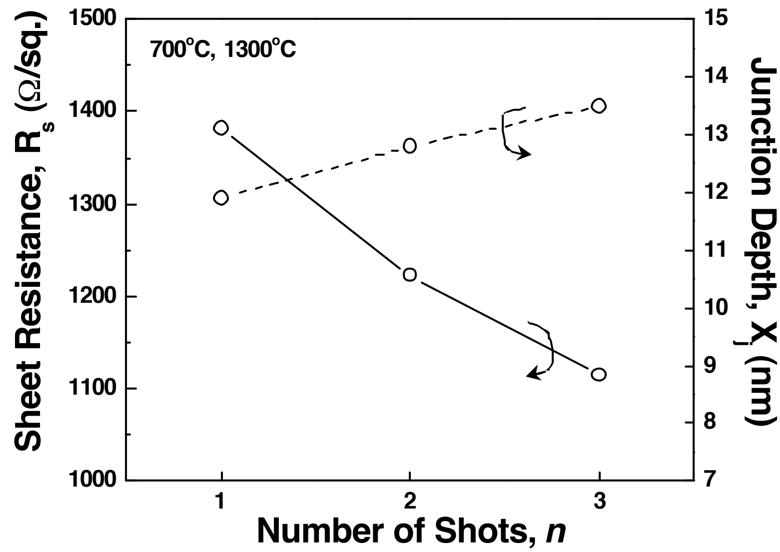


Figure 2.31 Measured sheet resistance and junction depth as a function of n for T_{int} = 700°C and T_{peak} = 1300°C.

The aforementioned results highlight how different flash anneal parameters affect junction depth and sheet resistance. For example, for a given T_{int} when T_{peak} is changed from 1250°C to 1300°C, X_j is increased and R_s is decreased. This reduction can be attributed to higher solid solubility limit of boron at 1300°C compared to 1250°C and/or to a relatively deeper junction. Similarly, when n is increased from 1 to 3, the associated reduction in R_s can be attributed to increased dopant activation and/or a change in X_j. It is very difficult to decouple these two possible causes by monitoring changes in R_s and X_j.

Therefore a simple model is presented here to compare the dopant activation efficiencies for different flash anneal conditions. The effective dopant activation efficiency, η, is defined as ?

$$\eta \equiv \frac{R_{SIMS}}{R_{meas}} \tag{2.2}$$

where R_{SIMS} is the sheet resistance calculated from the SIMS profile assuming 100% dopant activation, and R_{meas} is the measured sheet resistance of the same junction. R_{SIMS} is calculated by dividing the doped semiconductor into multiple parallel slices, as shown in Figure 2.32.

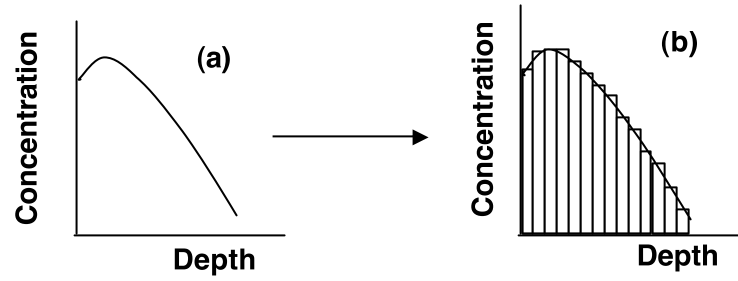


Figure 2.32 (a) Measured SIMS profile, **(b)** SIMS profile divided into multiple parallel slices.

Each slice of width Δx is assumed to be uniformly doped to a concentration $N(x)$, so that it has a sheet resistance

$$R_{s,j} = \frac{\left(qN(x)\mu\right)^{-1}}{\Delta x} \tag{2.3}$$

The sheet resistance of the entire junction is then obtained using the following equation:

$$\frac{1}{R_{SIMS}} = \sum_{j} \frac{1}{R_{s,j}} \tag{2.4}$$

The effective dopant activation efficiencies for different flash anneal conditions are summarized in Figure 2.33.

It is observed that η improves from 51% to 58% when T_{int} is reduced from 800°C to 700°C, respectively, for T_{peak} = 1250°C and n = 1. For a single-shot flash anneal, η is improved when T_{peak} is increased from 1250°C to 1300°C: 70% dopant activation efficiency is achieved for T_{int} = 700°C and T_{peak} = 1250°C, for n = 1. These findings are consistent with conclusions drawn from R_s and X_j measurements. The highest effective

dopant activation efficiency (η ~82%) is achieved with 3 flash anneal shots for T_{int} = 700°C and T_{peak} = 1300°C. The $R_s \times X_j$ product was also found to be minimized for T_{int} = 700°C, T_{peak} = 1300°C, n = 3.

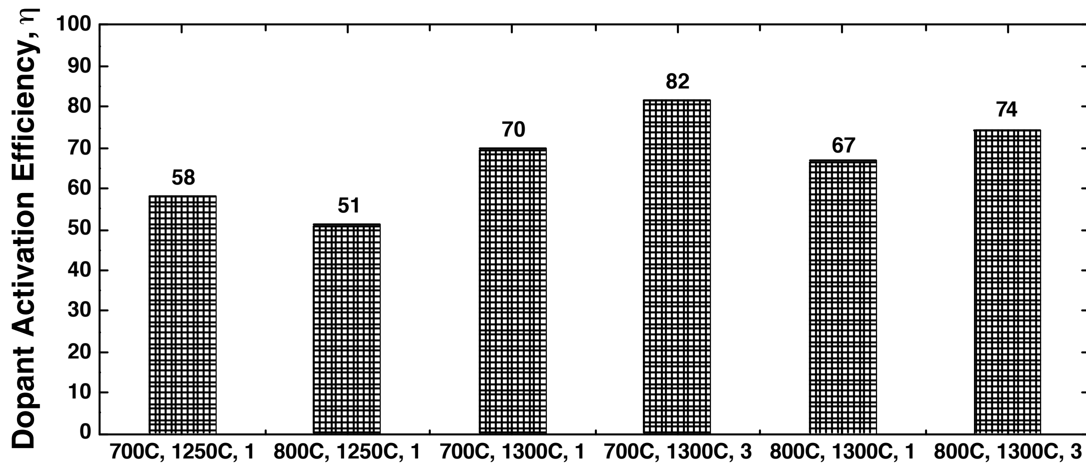


Figure 2.33 Effective dopant activation efficiencies for various flash anneal conditions.

This model does not give an accurate estimate of dopant activation level achieved by a given anneal condition, yet it is a useful tool to compare the efficacies of various anneals to achieve higher dopant activation (and thus lower R_s).

Figure 2.34 summarizes P+USJ experimental results, which show that it is possible to obtain X_j in the range ~12-15nm with R_s ~1kΩ/ via flash annealing, which is adequate for 32nm technology. Clearly, flash annealing not only reduces dopant diffusion but also improves dopant activation, as compared with spike annealing. N+USJ experimental results are shown in Figure 2.35 indicating that 32 nm node requirements are met for n-type junctions as well. 10nm deep junctions with ~1kΩ/ resistance are achieved.

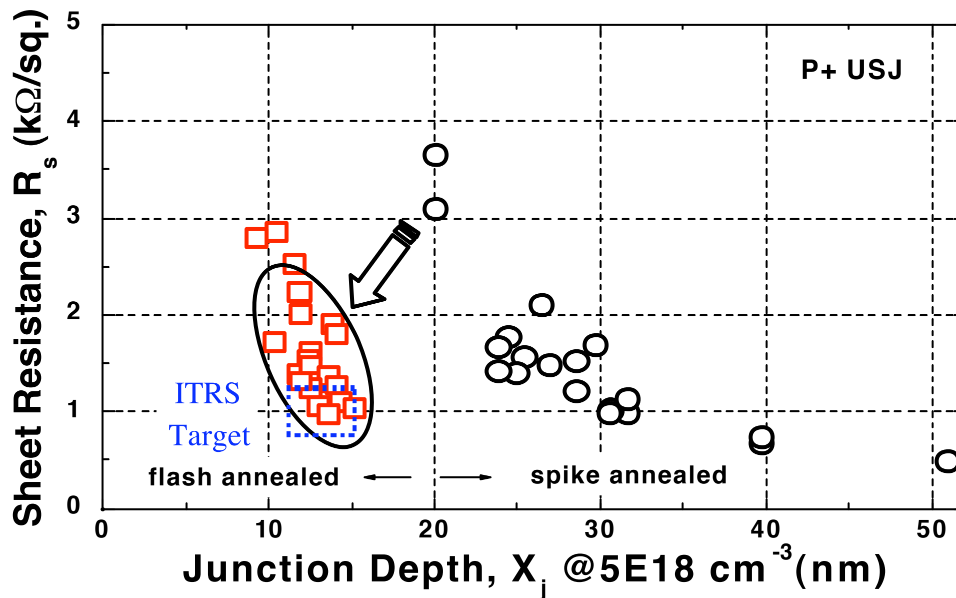


Figure 2.34 R_s-X_j summary $p+/n$ junctions formed by spike annealing (1070ºC) or flash annealing. Flash annealing provides for less dopant diffusion and more dopant activation.

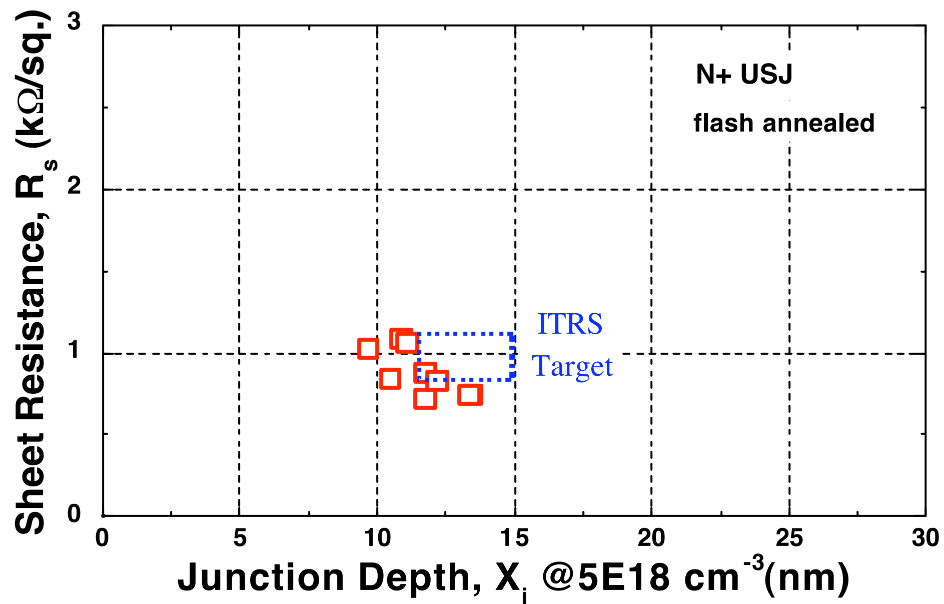


Figure 2.35 Sheet resistance (R_s) determined by non-contact method *vs*. junction depth (X_j) for $p+/n$ junctions formed by flash annealing.

One to the major challenges associated with advanced annealing technique is their efficacy to annihilate implantation damage because of the shorter anneal times. If the

residual defects lie within the junction depletion layer, it will cause excessive junction leakage due to carrier generation (recombination) when the *p*/*n* junction is reverse (forward) biased. High junction leakage may then become a significant contributor to total device leakage when the MOSFET is operating in the OFF state (V_{GS}=0, V_{DS} =V_{DD}). Figure 2.36 shows the junction leakage measured by non-contact method for various *p*+/*n* junctions that underwent either spike or flash anneal. PAI Ge implant was done prior to boron implantation for all the junctions and amorphized silicon depth is 36 nm. It is seen that flash-annealed junctions leak more compared to spike-annealed junction by ~100×. In all of the studied samples, the EOR damage band is located ~35 nm below the surface and down to 65nm as shown in high-resolution and weak beam dark field TEM images (Figure 2.37). This band is expected to lie within the depletion width of the junction, indicating that flash annealing is not effective for damage removal compared to spike annealing.

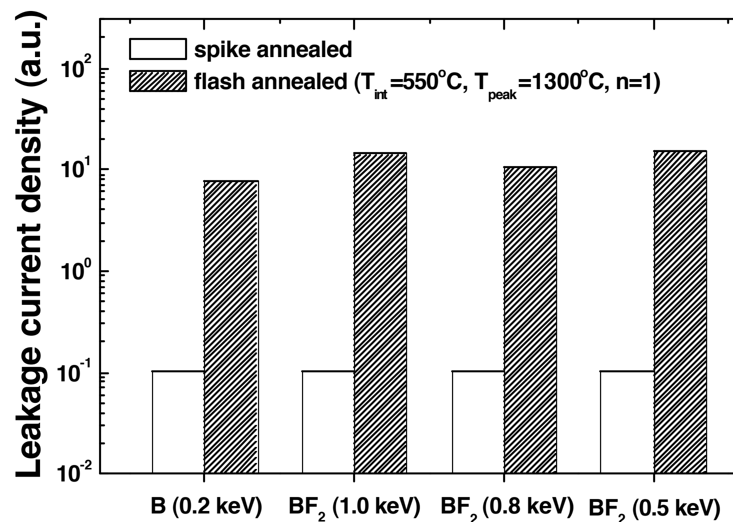


Figure 2.36 Junction leakage measured by non-contact method for various *p*/*n* junctions. Substrate was pre-amorphized prior to boron implantation.

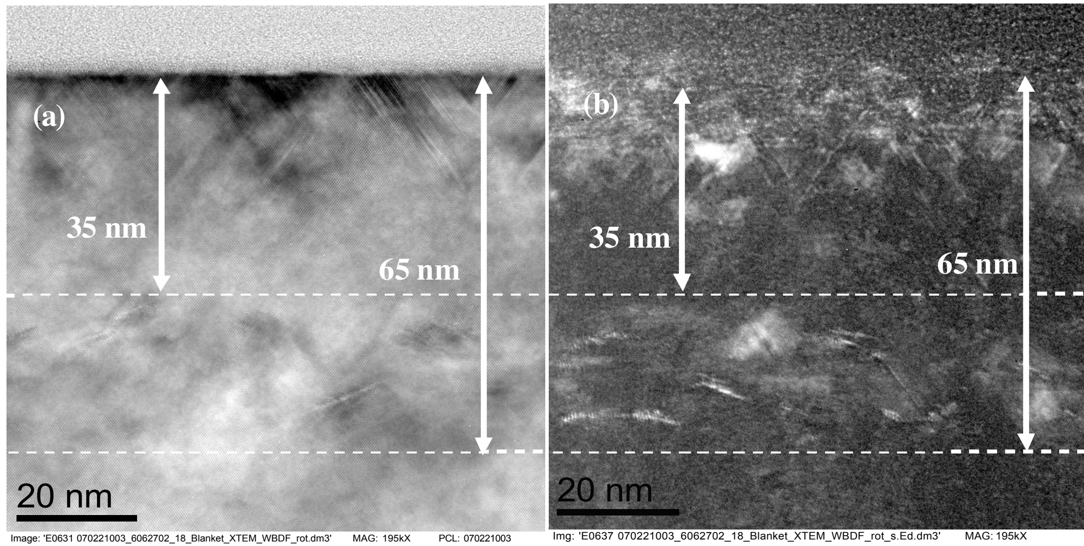


Figure 2.37 **(a)** Cross-sectional High resolution TEM micrograph — a high density of stacking faults and dislocations is observed near the surface and ~35 nm below the surface. **(b)** Weak beam dark filed (WBDF) TEM micrograph — WBDF g_{220} condition is used to study {311} defects. In the image, {311} defects are observed as wiggled inclined bright lines.

However with F (11 keV, 10^{15} cm^{-2}) co-implantation, junction leakage of flash-annealed samples can be reduced as shown in Figure 2.38. The implantation energy is picked so that most of the F dose lies in between boron and EOR damage band interface. 5-135× reduction in leakage is observed, depending upon the initial implanted boron profile. It is believed that F passivates the defects present in the EOR band and thus acts as an efficient sink or gettering site for these defects.

These results indicate the possibility of using F co-implantation in conjunction with flash annealing for reducing junction leakage while maintaining its benefits — shorter diffusion time and higher dopant activation.

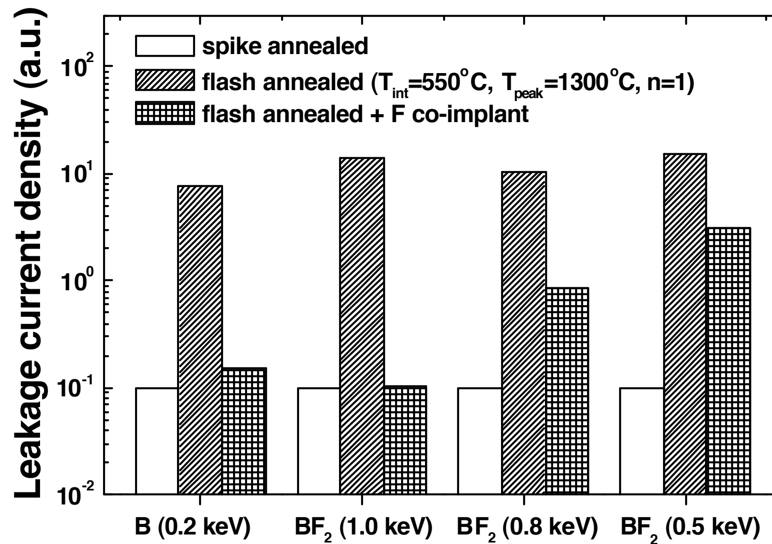


Figure 2.38 F co-implantation helps to reduce leakage in flash annealed *p*/*n* junctions.

2.8 Experimental Results with Infusion (GCIB) Doping

USJ experiments were performed on lightly doped (8-24 Ω-cm) *n*-type (100) and (110) silicon wafers. Using 1% B_2H_6/Ar as the source gas, boron was introduced by a 5kV or 3kV GCIB doping process to form *p*/*n* junctions. The GCIB dose was adjusted to give an equivalent 10^{15} cm^{-2} boron dose. A spike anneal (1070°C) or flash anneal was done to activate the introduced dopants. Flash anneal parameters (T_{int}, T_{peak}, and *n*) were varied to investigate their impact for GCIB doping process.

Figure 2.39 shows SIMS analyses of samples as-implanted with BF_2 (0.8keV, 2.5×10^{15} cm^{-2}, 7°) and as-infused with B (3kV, 10^{14} cm^{-2}). No pre-amorphization implant was done for the sample with BF_2 implantation. The as-infused profile indicates no evidence of channeling for (100) silicon substrate in contrast to the implanted profile, which shows a channeling tail starting at ~10^{19} cm^{-3}.

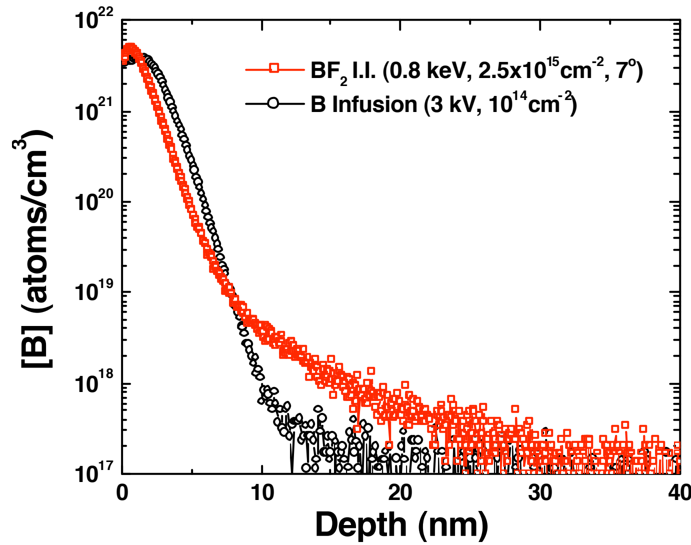


Figure 2.39 SIMS results for BF_2 implantation and B infusion processes before activation anneal. No PAI was done prior to BF_2 implantation.

Figure 2.40 shows spike and flash annealed profiles of for a 3kV, 10^{14} cm^{-2} B infusion process. The as-infused profile is also shown for comparison with, 5×10^{18} cm^{-3} level as a reference to measure junction depth. Spike annealing results in much deeper junctions (>40nm) whereas flash annealing causes little dopant movement as earlier observed for conventional ion implantation.

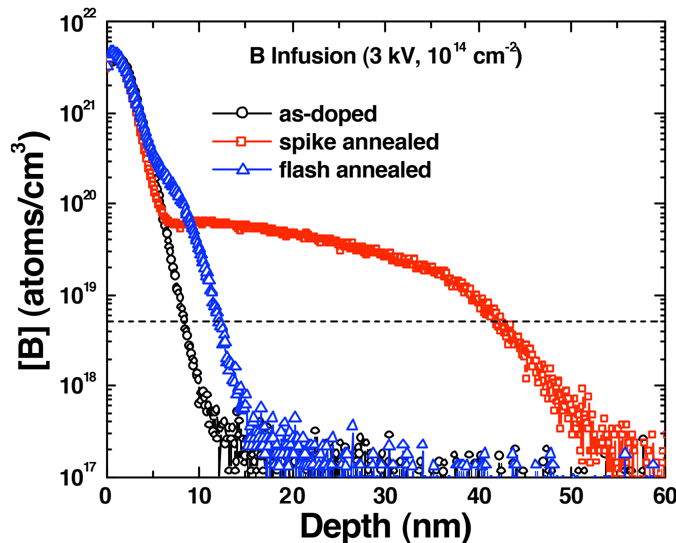


Figure 2.40 SIMS analyses of samples doped with B Infusion process (3kV, 10^{14} cm^{-2}). Negligible diffusion is observed for the flash annealed samples compared to the spike annealed samples.

Figures 2.41-2.43 summarize the effects of various flash anneal parameters investigated in this study. Figure 2.41 shows the impact of T_{int} for two different infusion process conditions. R_s is reduced considerably when T_{int} is changed from 800°C to 700°C for both 1250°C and 1300°C peak temperature. A small change in X_j is also observed, however this difference is within the depth resolution of SIMS. Thus, 700°C is preferred over 800°C for the intermediate temperature to achieve higher dopant activation.

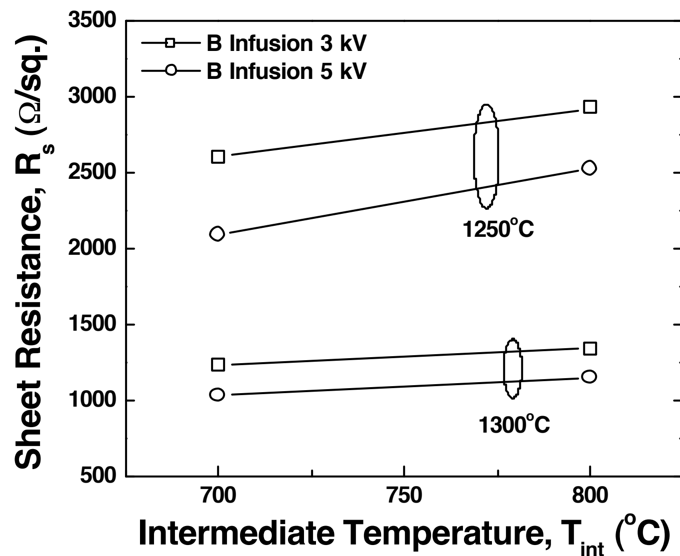


Figure 2.41 Measured sheet resistance as a function of T_{int} for two different peak temperatures.

Figure 2.42 illustrates the effect of T_{peak} for two different values of T_{int}. As seen from the reduction in R_s, T_{peak}=1300°C appears to be better choice for peak temperature. 1300°C also caused more dopant diffusion (ΔX_j ~0.5nm), but this change may be attributed to the SIMS depth resolution limitation. The combination of 700°C intermediate temperature with 1300°C as peak temperature is found to be optimal in this study. Figure 2.43 shows the impact of carrying out multiple shots of flash annealing. 40-

45% decrease in R_s is observed as the number of shots, n, is increased from 1 to 3. However, the junction depth is also increased (ΔX_j ~4.5nm).

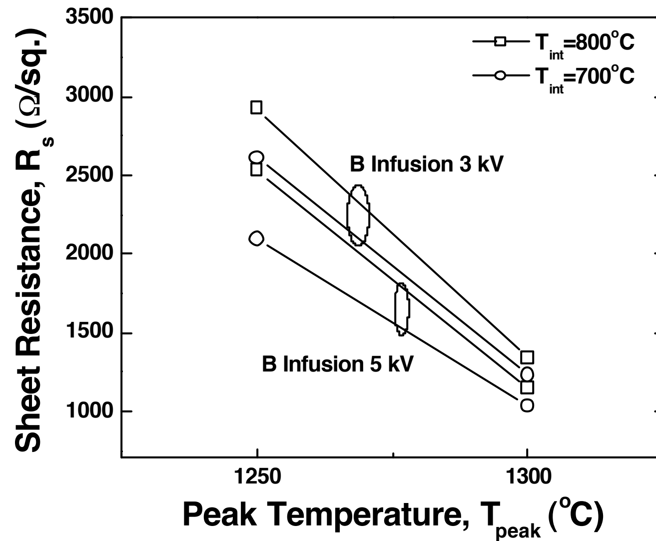


Figure 2.42 Measured sheet resistance as a function of T_{peak}.

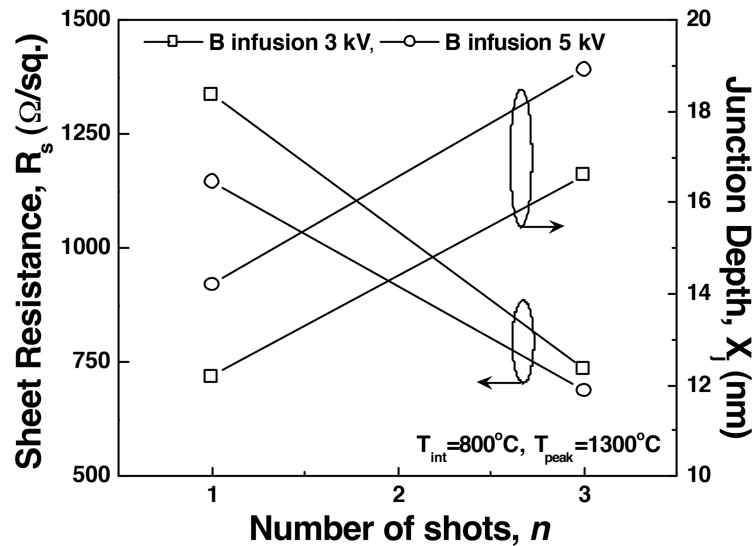


Figure 2.43 Measured sheet resistance and junction depth as a function of n for T_{int} = 800°C and T_{peak} = 1300°C.

R_s-X_j summary of flash annealed *p+/n* junctions formed by infusion doping is shown in Figure 2.44. Clearly, 32nm node USJ requirements are met (X_j~12-15nm, R_s~1kΩ/) with infusion doped, flash annealed junctions.

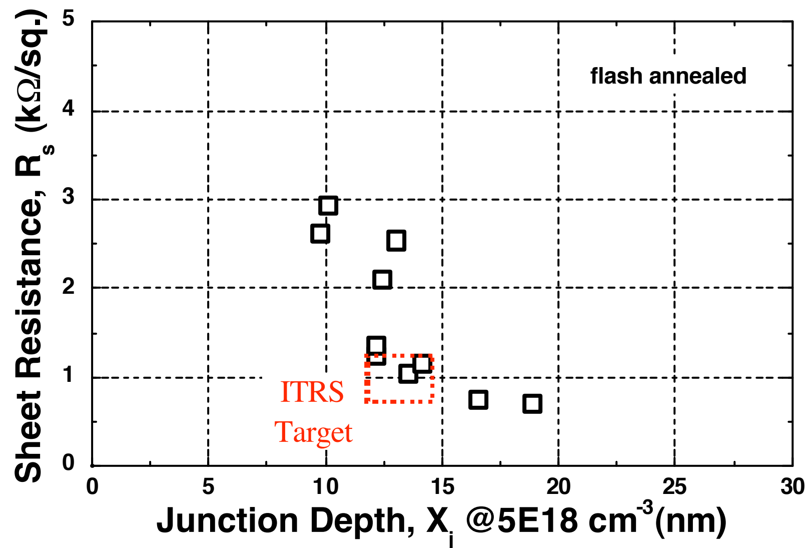


Figure 2.44 Sheet resistance (R_s) determined by non-contact method *vs.* junction depth (X_j) for *p+/n* junctions formed by infusion doping and flash annealing.

To keep pace with high drive current requirement for future CMOS generations, (110) pMOSFET has been proposed as a potential solution because of its higher hole mobility as compared with (100) oriented silicon [33]. However because of wider channels in (110) oriented crystal shown in Figure 2.45, ion channeling is a big concern as. This will result in large X_j making suppression of short channel effects difficult in (110) oriented pMOSFETs.

Infusion doping does not suffer from the problem of ion channeling as reported earlier in Figure 2.39. This feature of Infusion doping can be used to overcome the limitation of ion implantation. Figure 2.46 shows SIMS results obtained with (110) and (100) oriented silicon substrates before and after flash annealing. There appears to be

very small difference in the SIMS profiles obtained for both substrates, but this difference can be attributed to the SIMS depth resolution limitation. These results indicated that infusion doping is orientation independent and thus can be used for making ultra-shallow junctions for (110) pMOSFETs, if channeling is the limiting factor.

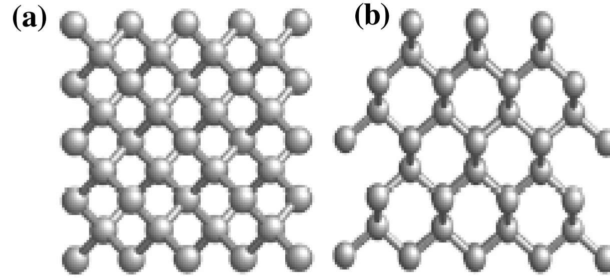


Figure 2.45 Atomic lattice of Si as viewed in the (a) <100> and (b) <110> directions.

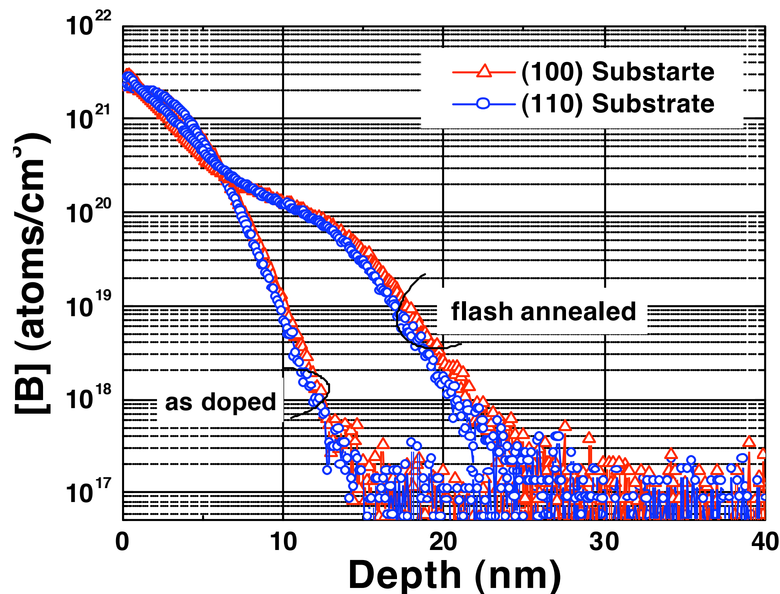


Figure 2.46 SIMS profiles of as doped and flash annealed (T_{int} = 800°C, T_{peak} = 1300°C, n=3) doped with boron infusion process (5kV, 10^{14} cm^{-2}) for (100) and (110) substrate orientations.

2.9 Summary

Ultra-shallow junction (USJ) requirements for sub-45 nm CMOS generations and associated challenges are reviewed in this chapter. Flash annealing appears to be a

promising candidate to replace conventional spike annealing. A detailed investigation is carried out to investigate the effects of flash anneal parameters. Best results are obtained with T_{int} = 700°C and T_{peak} = 1300°C. Multiple shots of flash annealing further lower R_s, with a trade-off in X_j. Non-contact method is employed for accurate R_s measurements. Furthermore, it is demonstrated that F co-implantation helps to reduce the junction leakage, so that the limitation of ms-anneals' efficacy to annihilate implant damage can be overcome with appropriate choice of F dose and energy.

Advantages of Infusion doping over ion implantation are discussed and it appears very promising for (110) oriented substrate. Because of lack of channeling for the infusion process, its incorporation in CMOS process flow will help to eliminate the need for PAI, and eliminate the associated EOR defects, which are formed upon annealing amorphized silicon. Hence, leakage of thusly formed junctions can be minimized.

This study indicates that to realize shallow junctions for future nanoscale CMOS nodes, the use of millisecond annealing will be imperative. However, the use of advanced doping techniques (e.g. plasma, infusion) may be delayed in high-volume manufacturing because of superior process control and in-depth understanding of ion implantation. Molecular implants will more likely be well suited for extending ion implantation for shallow doping applications.

2.10 References

[1] T. Skotnicki, "Heading for Decananometer CMOS – Is Navigation Among Icebergs Still a Viable Strategy?" *ESSDERC Proc.*, pp. 19-33, 2000.

[2] B. H. Lee, J. Oh, H. -H. Tseng, R. Jammy, and H. Huff, " Gate Stack Technology for Nanoscale Devices", *Materials Today*, vol. 9, no. 6, pp. 32-40, 2006.

[3] K. Mistry, C. Allen, C. Auth, B. Beattie, D. Bergstrom, M. Bost, M. Brazier, M. Buehler, A. Cappellani, R. Chau, C. -H. Choi, G. Ding, K. Fischer, T. Ghani, R. Grover, W. Han, D. Hanken, M. Hattendorf, J. He, J. Hicks, R. Huessner, D. Ingerly, P. Jain, R. James, L. Jong, S. Joshi, C. Kenyon, K. Kuhn, K. Lee, H. Liu, J. Maiz, B. Mclntyre, P. Moon, J. Neirynck, S. Pae, C. Parker, D. Parsons, C. Prasad, L. Pipes, M. Prince, P. Ranade, T. Reynolds, J. Sandford, L. Shifren, J. Sebastian, J. Seiple, D. Simon, S. Sivakumar, P. Smith, C. Thomas, T. Troeger, P. Vandervoorn, S. Williams, and K. Zawadzki, " A 45nm Logic Technology with High-k+Metal Gate Transistors, Strained Silicon, 9 Cu Interconnect Layers, 193nm Dry Patterning, and 100% Pb-free Packaging," *IEEE IEDM Tech. Digest*, pp. 247–250, 2007.

[4] International Technology Roadmap for Semiconductors, 2007 Edition, (http://www.public.itrs.net/).

[5] International Technology Roadmap for Semiconductors, 2005 Edition, (http://www.public.itrs.net/).

[6] P. A. Stolk, H. -J. Gossmann, D. J. Eaglesham, D. C. Jacobson, C. S. Rafferty, G. H. Gilmer, M. Jaraíz, J. M. Poate, H. S. Luftman, and T. E. Haynes, "Physical Mechanisms of Transient Enhanced Dopant Diffusion in Ion-Implanted Silicon," *J. Applied Physics*, vol. 81, no. 9, pp.6031-6050, 1997.

[7] P. -S. Chen, T. E. Hsieh, and C. -H. Chu, "Removal of End-of-range Defects in Ge^+-pre-amorphized Si by Carbon Ion Implantation," *J. Applied Physics*, vol. 85, no. 6, pp. 3114-3119, 1999.

[8] H. Graoui, M. Hilkene, B. McComb, M. Castle, S. Felch, N. E. B. Cowern, A. Al-Bayati, A. Tjandra, and M. A. Foad , "Optimization of Fluorine Co-implantation for PMOS Source and Drain Extension Formation for 65nm Technology Node," *Mat. Res. Soc. Symp. Proc.*, vol. 810, pp. 247-252, 2004.

[9] S. Nishikawa, A. Tanaka, and T. Yamaji, "Reduction of Transient Boron Diffusion in Preamorphized Si by Carbon Implantation," *Applied Physics Letters*, vol. 60, no. 18, pp. 2270-2272, 1992.

[10] N. E. B. Cowern, B. Colombeau, J. Benson, A. J. Smith, W. Lerch, S. Paul, T. Graf, F. Cristiano, X. Hebras, and D. Bolze, "Mechanisms of B deactivation control by F co-implantation", *Applied Physics Letters*, vol. 86, no. 10, pp. 101905, 2005.

[11] T. Murakami, T. Kuroi, Y. Kawasaki, M. Inuishi, Y. Matsui, and Akihiko Yasuoka, "Application of nitrogen implantation to ULSI," *Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms*, vol. 121, no. 1-4, pp. 257-261, 1997.

[12] S. B. Felch, B. S. Lee, S. L. Daryanani, D. F. Downey, and R. J. Matyi, "Characterization of Ultra-shallow p+/n junctions Formed by Plasma Doping with BF_3 and N_2 Plasmas," *Materials Chemistry and Physics*, vol. 53, pp. 37-43, 1998.

[13] M. Kase, "Ultra-Shallow Junction Formation," *IEEE VLSI Technology Short Course,* 2006.

[14] R. B. Liebert, S. R. Walther, S. B. Felch, Z. Fang, B. O. Pedersen, and D. Hacker, " Plasma doping system for 200 and 300 mm wafers", *Int. Conf. Ion Implantation Technology Proc.*, pp. 472 - 475, 2000.

[15] J. T. Scheuer, D. Lenoble, J.-P. Reynard, F. Lallement, A. Grouillet, A. Arevalo, D. Distaso, D. Downey, Z. Fang, L. Godet, B. W. Koo, T. Miller, and J. Weeman, "USJ formation using pulsed plasma doping," *Surface Coatings & Technology*, vol. 186, pp. 57-61, 2004.

[16] J. Hautala, J. Borland, M. Tabat, and W. Skinner, "Infusion Doping for USJ Formation," *Int. Workshop Junction Technology*, pp. 50-53, 2004.

[17] J. Hautala, M. Gwinn, W. Skinner, and Y. Shao, "Productivity Enhancements for Shallow Junctions and DRAM Applications using Infusion Doping," *Int. Conf. Ion Implantation Technology Proc.*, pp. 174-177, 2006.

[18] W. Krull, B. Haslam, T. Horsky, K. Verheyden, and K. Funk, " Simplifying the 45nm SDE Process with ClusterBoron® and ClusterCarbon™ Implantation," *Int. Conf. Ion Implantation Technology Proc.*, pp. 182-185, 2006.

[19] L. M. Rubin, M. S. Ameen, M. A. Harris, and H. Chuong, "Molecular Implants for Advanced Devices," *Int. Workshop Junction Technology*, pp. 113-118, 2007.

[20] T. Ito, T. Iinuma, A. Murakoshi, H. Akutsu, K. Suguro, T. Arikado, K. Okumura, M. Yoshioka, T. Owada, Y. Imaoka, H. Murayama, and T. Kusuda, "Flash Lamp Anneal Technology for Effectively Activating Ion Implanted Si," *Ext. Abs. SSDM*, pp. 182–183, 2001

[21] T. Gebel, M. Voelskow, W. Skorupa, G. Mannino, V. Priviter, F. Priolo, E. Napolitani, and A. Carnera, "Flash Lamp Annealing with Millisecond Pulses for Ultra-shallow Boron Profiles in Silicon," *Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms*, vol. 186, no. 1-4, pp. 287-291, 2002.

[22] B. J. Cho, D. Poon, L. S. Tan, M. Bhat and A. See, "Multiple-Pulse Laser Annealing of Boron-Implanted Preamorphized Silicon and the Process Optimization," *Int. Workshop Junction Technology*, pp. 22-26, 2004.

[23] R. Lindsay, K. Henson, W. Vandervorst, K. Maex, B. J. Pawlak, R. Duffy, R. Surdeanu, P. Stolk, J. A. Kittla, S. Giangrandi, X. Pages, and K. van der Jeugd, " Leakage Optimization of Ultra-shallow Junctions Formed by Solid Phase Epitaxial Regrowth," *J. Vac. Sci. Technol. B*, vol. 22, no. 1, pp. 306-311, 2004.

[24] T. Clarysse, D. Vanhaeren, and W. Vandervorst, "Impact of Probe Penetration on the Electrical Characterization of Sub-50 nm Profiles," *J. Vac. Sci. Technol. B*, vol. 20, no. 1, pp. 459-466, 2002.

[25] V. N. Faifer, M. I. Current, T. Nguyen, T. M. H. Wong, V. V. Souchkov, "Non-contact Measurement of Sheet Resistance and Leakage Current: Applications for USJ-SDE/Halo Junctions," *Int. Workshop Junction Technology*, pp. 45-48, 2005.

[26] V. N. Faifer, M. I. Current, T. M. H. Wong, and V. V. Souchkov, "Noncontact sheet resistance and leakage current mapping for ultra-shallow junctions," *J. Vac. Sci. Technol. B*, vol. 24, no. 1, pp. 414-420, 2006.

[27] V. N. Faifer, M. I. Current, W. Walecki, V. Souchkov, G. Mikhaylov, P. Van, T. Wong, T. Nguyen, J. Lu, S. H. Lau, and A. Koo, "Non-Contact Electrical Measurements of Sheet Resistance and Leakage Current Density for Ultra-Shallow (and Other) Junctions," *Mat. Res. Soc. Symp. Proc.*, vol. 810, pp. 475-480, 2004.

[28] C. H. Poon, B. J. Cho, Y. F. Lu, M. Bhat, and A. See, "Multiple-pulse Laser Annealing of Preamorphized Silicon for Ultrashallow Boron Junction Formation," *J. Vac. Sci. Technol. B*, vol. 21, no. 2, pp. 706-709, 2003.

[29] A. Shima, H. Ashihara, T. Mine, Y. Goto, M. Horiuchi, Y. Wang, S. Talwar and A. Hiraiwa, "Self-limiting Laser Thermal Process for Ultra-shallow Junction Formation of 50-nm Gate CMOS," *IEEE IEDM Tech. Digest*, pp. 493–496, 2003.

[30] P. Timans, J. Gelpey, S. McCoy, W. Lerch, and S. Paul, "Millisecond Annealing: Past, Present and Future," *Mat. Res. Soc. Symp. Proc.*, vol. 912, pp. 3-14, 2006.

[31] T. Ito, K. Suguro, T. Itani, K. Nishinohm, K. Matsuo, and T. Saito, "Improvement of Threshold Voltage Roll-of by Ultra-Shallow Junction Formed by Flash Lamp Annealing," *IEEE VLSI Technology Symp. Tech. Digest*, pp. 53-54, 2003.

[32] K. Adachi, K. Ohuchi, N. Aoki, H. Tusjii, T. Ito, H. Itokawa, K. Matsuo, K. Suguro, Y. Honguh, N. Tamaoki, K. Ishimaru, and H. Ishiuchi, "Issues and Optimization of Milisecond Anneal Process for 45 nm node and Beyond," *IEEE VLSI Technology Symp. Tech. Digest*, pp. 142-143, 2005.

[33] M. Yang, M. Ieong, L. Shi, K. Chan, V. Chan, A. Chou, E. Gusev, K. Jenkins, D. Boyd, Y. Ninomiya, D. Pendleton, Y. Surpris, D. Heenan, J. Ott, K. Guarini, C. D'Emic, M. Cobb, P. Mooney, B. To, N. Rovedo, J. Benedict, R. Mo, and H. Ng, "High Performance CMOS Fabricated on Hybrid Substrate with Different Crystal Orientations," *IEEE IEDM Tech. Digest*, pp. 453-456, 2003.

Chapter 3

Process Integration of High-*k*/Metal-Gate with Flash Anneal

3.1 Introduction

Poly-Si gate and SiO_2 gate dielectric have been used as the standard gate stack in CMOS technologies over the years, until recently. The use of thermally grown SiO_2 offers a thermally and electrically stable high-quality Si/SiO_2 interface. Indeed, the key enabler of successful MOSFET scaling has been the thinning of SiO_2 with every generation for increased gate control. The gate dielectric thickness has shrunk from ~100 nm in early generations to a few atomics layers (~1-2 nm) in current technology nodes. However, as SiO_2 is scaled below 2 nm in thickness, it starts to conduct a significant amount of current due to the quantum mechanical tunneling of electrons or holes, as shown in Figure 3.1 [1]. High gate leakage increases power consumption and dielectric reliability becomes a major concern for such an ultra-thin SiO_2 gate oxide. Furthermore,

boron can diffuse rapidly from a poly-Si gate to channel through a sub-2 nm oxide during high-temperature process steps, causing the threshold voltage to shift.

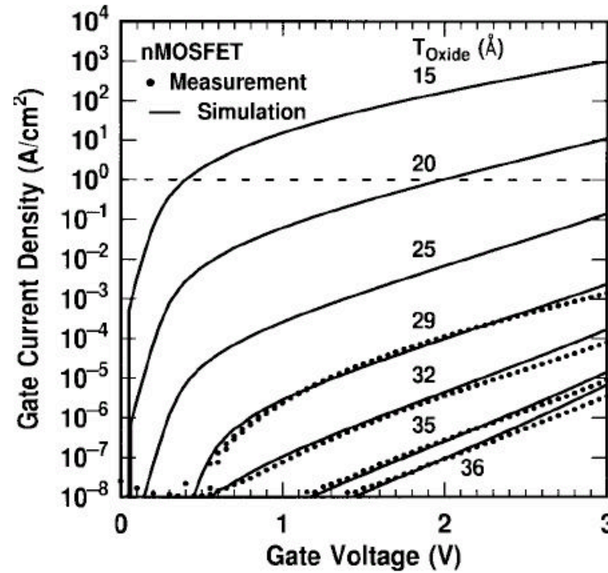


Figure 3.1 Measured and simulated direct tunneling gate current of nMOSFETs with ultra-thin SiO_2 dielectric [1].

The solution for avoiding undesired effects of ultra-thin gate dielectric is to replace SiO_2 with higher permittivity material. The equivalent oxide thickness (EOT) of a dielectric with relative permittivity κ and thickness t_{phys} is given as –

$$EOT = \frac{3.9}{\kappa} t_{phys} \tag{3.1}$$

where 3.9 is the relative permittivity for SiO_2. From a device operation point of view, a physically thicker higher-κ gate dielectric is equivalent to SiO_2 with a thickness given by the *EOT* value.

The addition of N to SiO_2 reduces boron diffusion through the dielectric and also provides for reduced leakage because the dielectric can be made slightly thicker for a given EOT. However, nitrided oxide provides only a near-term solution and its scaling is

limited to ~1.3 nm [2]. An alternate dielectric material with higher κ will be needed to replace silicon oxide/nitrided oxide for a long-term solution.

Another major problem with a conventional poly-Si/SiO_2 gate stack is the poly-Si depletion near the gate dielectric interface. The result is a ~3-4 Å thick parasitic dielectric layer which adds substantially to EOT for sub-2 nm gate stacks. Metal gates offer a potential solution to the poly-Si depletion problem. The current status of replacing the conventional gate stack with an advanced gate stack is discussed in the next section.

3.2 Advanced Gate Stacks

Many materials properties need to be considered to determine the best high-κ candidate to replace SiO_2. The material must offer a significant barrier against carrier tunneling because leakage current increases exponentially with decreasing barrier height. Thus, high-κ material must have sufficient conduction- and valence-band offsets with respect to the silicon substrate and gate material. Additionally, high-κ material must be thermodynamically stable on silicon, must provide high-quality interface with silicon channel, and must remain in an amorphous state (to minimize leakage current) at high temperatures commonly used in CMOS processing.

Several high-κ materials have been investigated as a gate dielectric to replace SiO_2, such as TiO_2 [3-5], Ta_2O_5 [6-8], La_2O_3 [9], Y_2O_3 [10], $SrTiO_3$ [11], Al_2O_3 [12-14], HfO_2 [15, 16], and ZrO_2 [17-19]. Table 3.1 lists the most commonly studied high-κ candidates and their corresponding dielectric constants, bandgaps (E_g), conduction band offsets (ΔE_c), merits, and drawbacks [20]. TiO_2, Ta_2O_5, and $SrTiO_3$ have very low ΔE_c and are not thermodynamically stable on silicon. In addition, TiO_2 has a low

crystallization temperature (~400°C). Y_2O_3 and La_2O_3 have large E_g and ΔE_c, but Y_2O_3 crystallizes at ~400°C, and chemical instability due to reaction with moisture is a serious issue for La_2O_3. Al_2O_3 has many favorable properties including a large E_g, and is amorphous and thermodynamically stable on silicon up to high temperatures. However, Al_2O_3 has κ value of only ~10 and has a large fixed charge density. Thus, Al_2O_3 will not be a viable candidate to replace SiO_2 as a long-term solution. ZrO_2 is a potentially attractive candidate, but it forms detrimental silicide when in contact with poly-Si gate.

Table 3.1 Comparison of various characteristics and main features of existing and candidate high-κ gate dielectrics [20].

Dielectric	Dielectric constant (bulk)	Bandgap (eV)	Conduction band offset (eV)	Merits	Drawbacks
Silicon dioxide (SiO_2)	3.9	8.9	3.15	Excellent Si interface, low Q_{ox} and D_{it}	Low-κ, EOT > 0.8 nm
Silicon nitride (Si_3N_4)	7–7.8	5.3	2.1	Good interface and bulk properties, medium Q_{ox} and D_{it}	Low-κ, EOT > 0.5 nm
Aluminum oxide (Al_2O_3)	9–10	8.8		E_g comparable to SiO_2, amorphous Good thermal stability	Medium Q_{ox} and D_{it}, medium κ
Tantulum pentoxide (Ta_2O_5)	25	4.4	0.36	High-κ	Unacceptable ΔE_C, not stable on Si,
Lanthana (La_2O_3)	~27	5.8	2.3	High-κ, better thermal stability	Moisture absorption, instable with Si
Gadolinium oxide (Gd_2O_3)	~12	~5	–[a]	Low D_{it} –[a]	High Q_{ox}, Crystallization
Yttrium oxide (Y_2O_3)	~15	6	2.3	Large E_g	Low crystallization temperatature, hight D_{it}, silicide formation
Hafnia (HfO_2)	~20	5.6–5.7	1.3–1.5	Most suitable compared to other candidates	Crystallization, silicate and silicide formation,
Zirconia (ZrO_2)	~23	4.7–5.7	0.8–1.4	Similar to hafnia	High Q_{ox} and D_{it} Marginal stable with Si, crystallization, silicide formation
Strontium titanate ($SrTiO_3$)	~300	3.3	–0.1	High-κ	Unacceptable E_g and ΔE_C, field fringing effect

Research on high-κ dielectrics has converged on HfO_2-based dielectrics because they show much promise in overall materials properties to replace SiO_2. However, typical dopant activation anneal temperatures (~1000°C) used in CMOS processing are much higher than the crystallization temperature of HfO_2. The addition of Si [21] and N [22] to HfO_2 increases thermal stability significantly. Improved thermal stability also minimizes the interfacial layer thickness. However, N induces a fixed oxide charge and tends to pile

up at the interface, which can lead to lower mobility. Furthermore, the addition of Si reduces the dielectric constant of the gate dielectric. Therefore, careful optimization of Si and N content and profile is needed to fully realize the scaling benefits of HfO_2.

Despite the growing understanding of Hf-based dielectrics, potential roadblocks remain for the successful integration of poly-Si and Hf-based dielectrics. An unfavorable increase in threshold voltage has been observed for PMOS devices with poly-Si/high-κ stack [23]. This is primarily due to Fermi-level pinning because the gate work function for a poly-Si/high-κ gate stack is determined by Si-Hf bonding instead of the Fermi level of the poly-Si gate. The pinning location is believed to be just below the Si conduction band edge [24]. Additionally, the poly-Si electrode suffers from a number of other drawbacks such as gate depletion and boron penetration. Therefore, research on the use of a metal gate electrode to replace conventional poly-Si has gained momentum. A HfSiON gate dielectric with metal gate has been shown to meet 45 nm node requirements with EOT scalable to 1 nm and carrier mobilities comparable to that of nitrided oxide [25]. Intel has implemented high-κ/metal-gate technology in 45 nm CMOS generation using a high-κ first, gate-last process integration approach [26].

It is well recognized that one of the major issues for nano-scale MOSFET source/drain design is the formation of ultra-shallow junctions (USJs). In chapter 2, it is shown that to achieve USJ requirements for the 32 nm generation and beyond, advanced anneal techniques such as flash anneal will be needed. Since high-κ/metal-gate technology is already being implemented at the 45 nm node, it becomes very important to understand the integration challenges of flash anneal with a high-κ/metal-gate. There are reports of using laser anneal and flash anneal on poly-Si/silicon oxide-based gate stack,

but very little is reported about their impact on high-κ/metal gate stacks [27-32]. In this chapter, a detailed investigation of the effects of flash annealing on MOSFETs with Hf-based dielectric and metal gate is presented.

3.3 Device Fabrication

A conventional gate-first CMOS process flow was used to fabricate sub-100 nm gate length devices on *p*-type (100) silicon wafers with a 5 μm epi-layer of doping density ~3×10^{17} cm^{-3} as shown in Figure 3.2. After a shallow trench isolation (STI) process, *n*-wells and *p*-wells were formed for *p*-MOS and *n*-MOS devices respectively. After a pre-gate clean and an ozone treatment that produces ~1 nm interfacial SiO_x, a 3-nm thick HfSiO film was grown by atomic layer deposition using pulses of tetrakis(ethylmethylamino)hafnium (TEMAHf) and tetrakis(ethylmethylamino)silicon (TEMASi) [33]. Ozone was the oxidizing ambient. The composition was determined to be Si/(Si+Hf)~0.20. Afterwards, wafers received a post-high-κ deposition anneal (PDA) in ammonia (NH_3) at 700°C and 30 Torr for 60 seconds. Resultant N content in HfSiON was estimated to be ~7 at. % by X-ray photoelectron spectroscopy (XPS) [34]. For the gate electrode, 10 nm TiN was deposited by ALD, and capped by 100 nm thick LPCVD poly-Si for ease of process integration. S/D extensions (SDE) and halo implants were done next, followed by spacer and deep S/D formation. For *n*-SDE, As was implanted at 0.5keV, 10^{15} cm^{-2} with 7° tilt. A pre-amorphization implant (Ge: 10 keV, 10^{15} cm^{-2}) followed by B implant (0.8 keV, 5×10^{14} cm^{-2}, 7°) were used to form *p*-SDE. A spike anneal (1070°C) or flash anneal (T_{int}=800°C, T_{peak}=1300°C) was used to activate implanted source/drain dopants, followed by a two-step cobalt-salicidation process done

at 550°C and 750°C for 30 seconds in N_2 ambient. After metallization, forming gas anneal (FGA) was done last in H_2/N_2 at 480°C for 30 minutes.

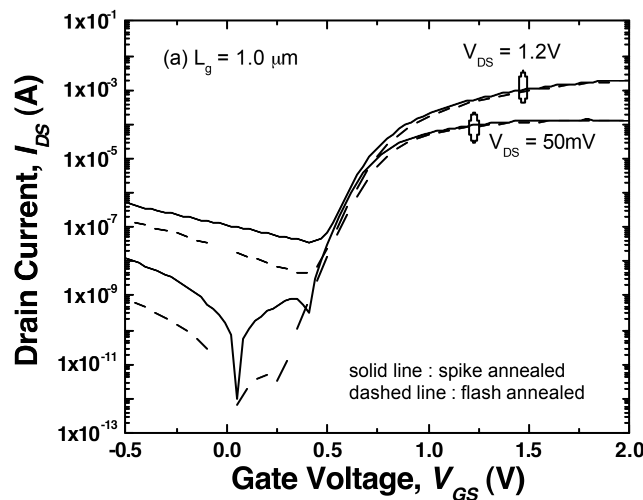
- **STI and Well Formation**
- **Gate Stack Formation**
  (ALD HfSiO + ALD TiN)
- **Shallow Extensions and Halo I/I**
- **Spacer Formation**
- **Deep Source/Drain I/I**
- **Dopant Activation Anneal**
  Spike (1070°C) *vs.* Flash (1300°C, ms)
- **Salicidation**
- **Metallization**

Figure 3.2 Gate-first CMOS process flow used to evaluate the impact of flash annealing on high-κ/metal-gate stacks.

3.4 Device Characterization

3.4.1 SCE Control

Well-behaved transistor characteristics are obtained from the fabricated MOSFETs. Figure 3.3 shows the measured I_{DS}-V_{GS} characteristics of spike annealed and lash annealed MOSFETs (L_g = 1.0 μm & 0.1 μm) at V_{DS} =50 mV and 1.2 V.



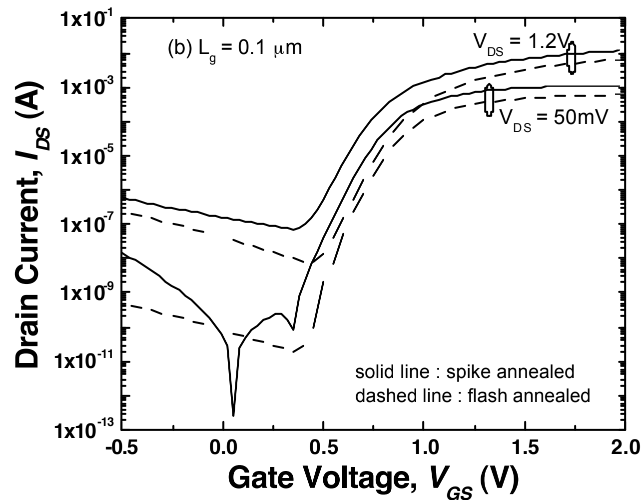


Figure 3.3 I_d-V_g of spike annealed and flash annealed nMOSFETs ? (a) long-channel (L_g = 1.0 μm) and (b) short-channel (L_g = 0.1 μm) ? measured at V_d = 50 mV and V_d = 1.2 V.

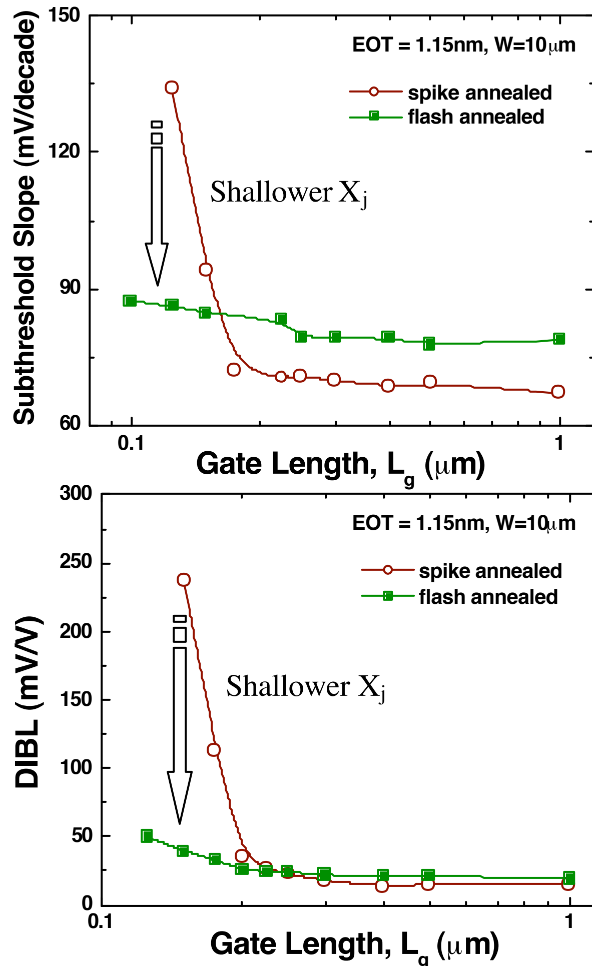


Figure 3.4 Measured SS and DIBL values for nMOSFETs fabricated using either spike or flash annealing to activate the implanted S/D dopants.

The subthreshold slope (SS) and drain-induced barrier lowering (DIBL) of the transistors with different gate lengths are shown in Figure 3.4. It can be seen that the short-channel effects (SCE) are well controlled down to 100 nm gate length for flash annealed MOSFETs as compared with spike annealed MOSFETs. The improved SCE can be attributed to shallower junction depth (X_j) achieved in flash annealed devices for identical source/drain-extension ion implantation conditions. However, SS for MOSFETs with L_g>0.15 µm is worse for flash annealed devices as compared with spike annealed devices. This observation is explained in section 3.4.3.

3.4.2 Gate Stack Integrity

The physical structure of gate stack after S/D activation anneal is analyzed using transmission electron spectroscopy (TEM). Figure 3.5 shows cross-sectional bright field TEM (BFTEM) and high resolution TEM (HRTEM) images for spike annealed and flash annealed gate stacks. These images illustrate that there is negligible difference in the physical thicknesses of HfSiON film and interfacial oxide layer for spike *vs.* flash annealed gate stacks. Measured thicknesses for HfSiON and SiO_x are ~3 nm and ~1 nm, respectively. The HfSiON layer is observed to be amorphous after spike (or flash) anneal, based on the evidence provided by the HRTEM images. The HfSiON/SiO_x interface appears to be slightly rougher in the flash annealed sample.

High angle annular dark-field scanning Transmission Electron Microscopy (HAADF-STEM) using energy-dispersive X-ray spectroscopy (EDXS) and electron energy loss spectra (EELS) is used to compare chemical profiles after S/D activation anneal as shown in Figure 3.6. EELS and EDXS data were acquired across the high-κ/ metal-gate stack. EDXS and EELS profiles relate data with different intensity scales,

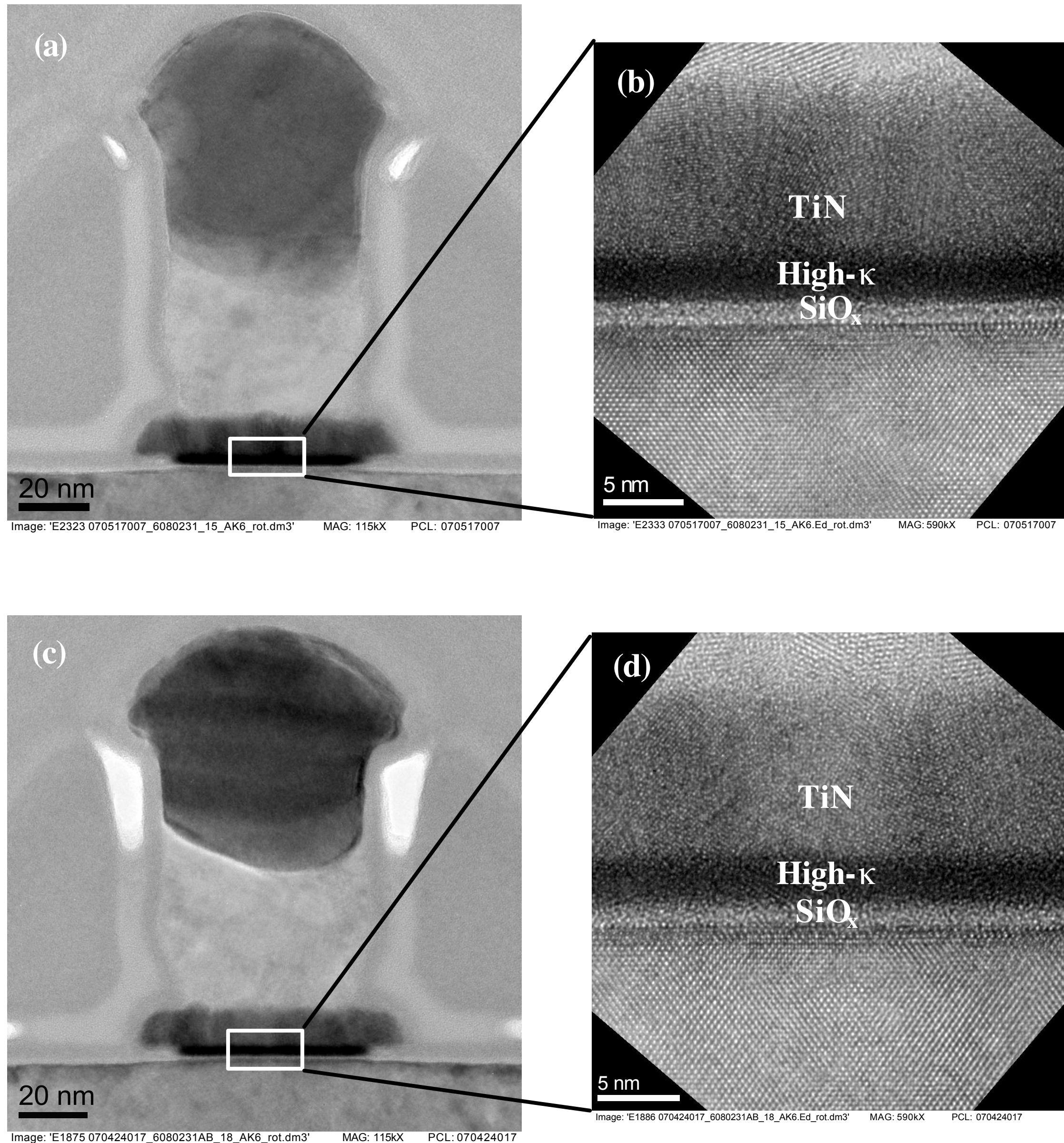


Figure 3.5 Cross-sectional bright field TEM (BFTEM) and high resolution TEM (HRTEM) images of high-κ/metal gate stacks after S/D activation anneal. (a), (c) ? BFTEM; (b), (d) ? HRTEM.
(a), (b) ? spike anneal (1070°C); (c), (d) ? flash anneal (T_{int}=800°C, T_{peak}=1300°C) The thickness of the TiN, the HfSiON film and the interfacial oxide layer are approximately 9.7 nm, 3 nm and 0.9 nm, respectively, for both spike and flash annealed devices.. The HfSiON layer was observed to be amorphous, based on the evidence provided by the HRTEM images.

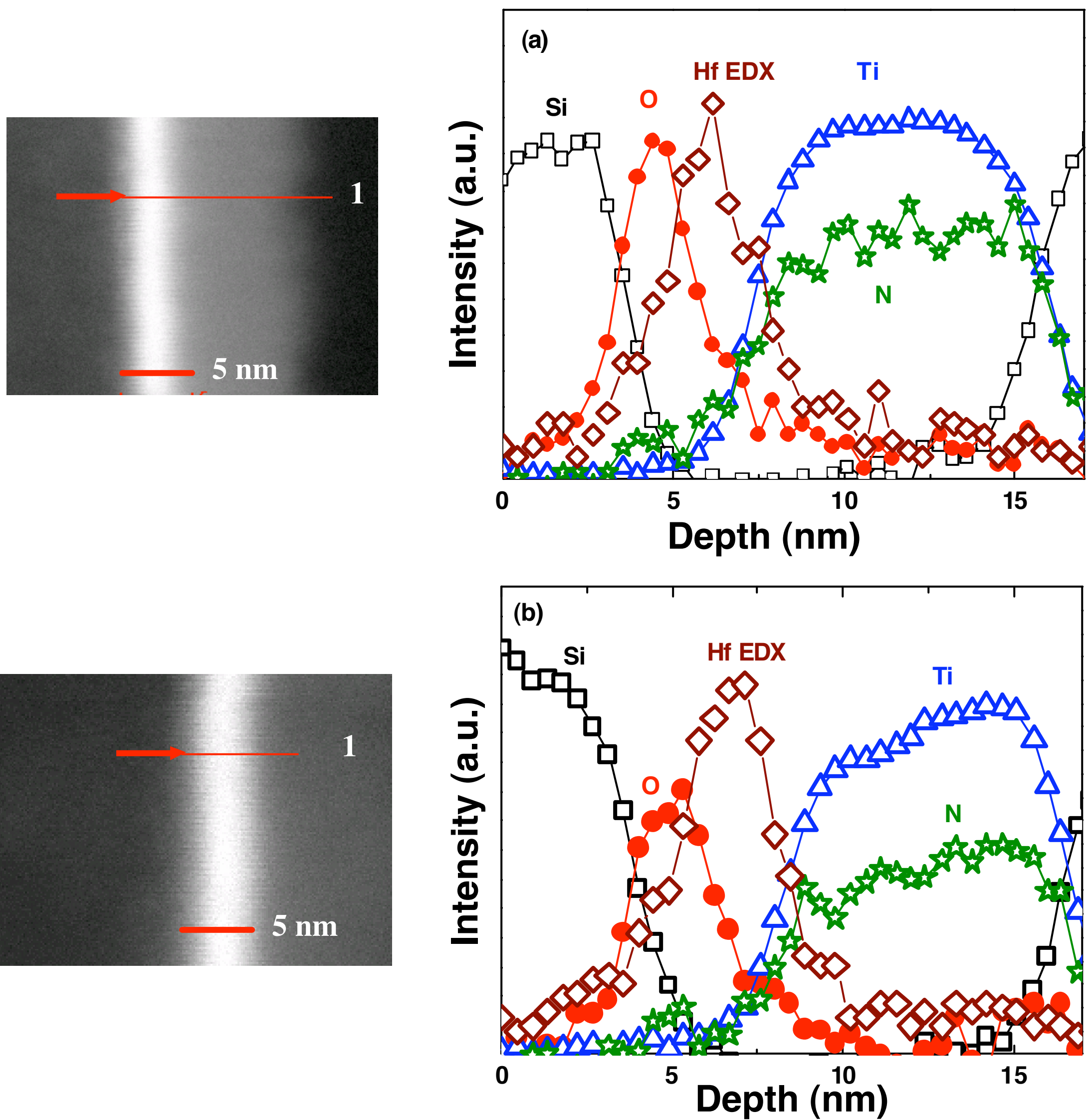


Figure 3.6 High angle annular dark-field scanning TEM (HAADF-STEM) micrograph of the sample in cross-section with corresponding electron energy loss spectrometry (EELS) and energy dispersive X-ray spectroscopy (EDXS) scans acquired along the line from left to right as indicated by the arrow.
(a) ? spike anneal (1070°C); (b) ? flash anneal (T_{int}=800°C, T_{peak}=1300°C)

but have been scaled to allow comparison on one plot. Figure 3.6 suggests that there is no noticeable difference in the chemical profiles for gate stacks which underwent flash annealing *vs*. spike annealing.

Capacitance-voltage (C-V) measurements were done next on high-κ/metal-gate MOS capacitors (A=5×10^{-5} cm^2). Figure 3.7 shows the well-behaved high-frequency (f = 100 kHz) C-V curve for nMOS capacitors. The equivalent oxide thickness (EOT) and the flatband voltage (V_{FB}) were extracted using the NCSU CVC program [35]. Comparable extracted EOT (~1.15 nm) and V_{FB} (~-0.52 V) values are seen for the gate stacks exposed to flash *vs*. spike anneals. pMOS capacitors exhibit similar features as shown in Figure 3.8 and negligible difference in extracted EOT and V_{FB} is observed.

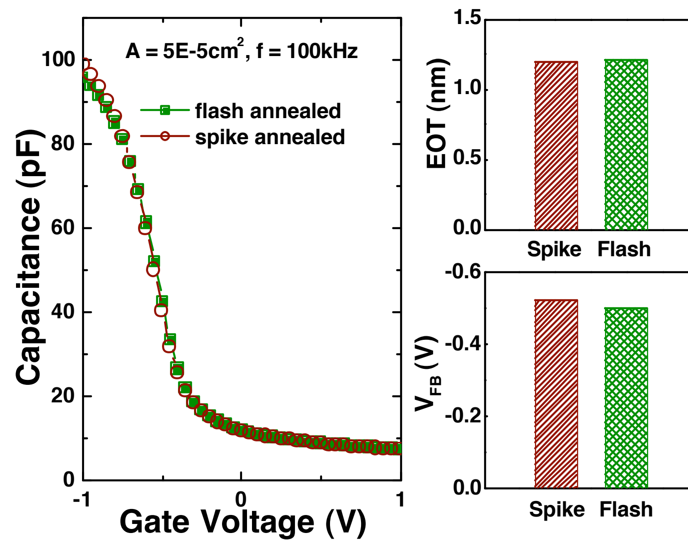


Figure 3.7 Comparison of C-V curves for nMOS capacitors fabricated with spike or flash annealing (f = 100 kHz). Comparable EOT and V_{FB} values are seen (right).

As shown in Figure 3.9, no frequency dispersion is seen in the multi-frequency C-V measurements for the flash annealed gate stacks as frequency is increased from 10 kHz to 250 kHz. The C-V hysteresis measured using bidirectional voltage sweeps is negligible at a measurement frequency of 100 kHz.

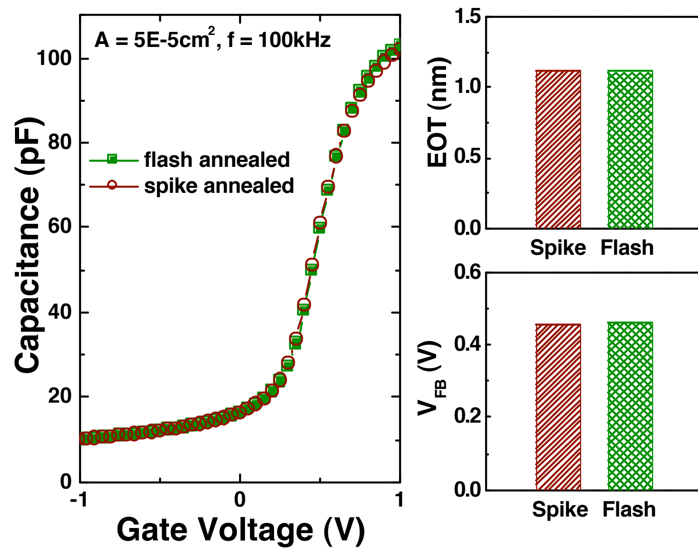


Figure 3.8 Measured high-frequency (f = 100 kHz) C-V curves for high-κ/metal-gate pMOS capacitors after either spike or flash annealing. Similar EOT (~1.15 nm) and V_{FB} (~0.45 V) values are achieved as shown on the right.

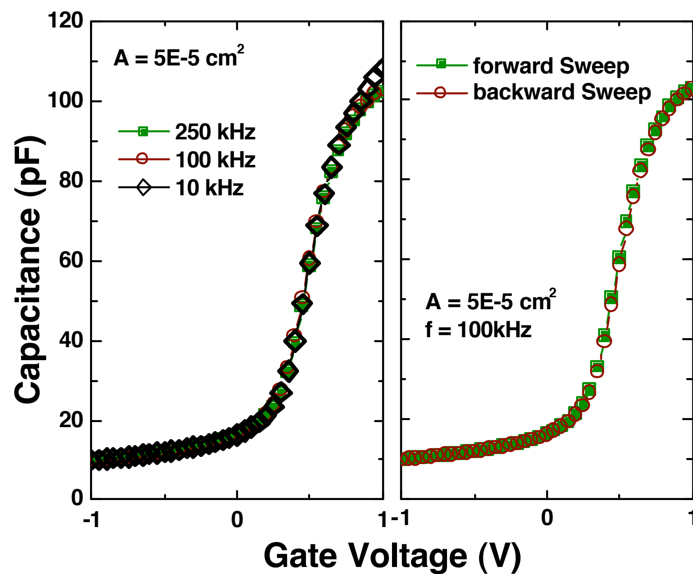


Figure 3.9 C-V curves for pMOS capacitors fabricated with flash annealing. No frequency dispersion or hysteresis is observed for flash annealed gate stacks.

The accumulation gate leakage current of capacitors (A=5×10^{-5} cm^2) fabricated with spike or flash annealing is shown in Figure 3.10. The leakage current of the flash annealed gate stack is comparable to that of the spike annealed gate stack at a bias of V_{GS}-

V_{FB}=1 V (J_g = 0.5 A/cm^2). The time-zero breakdown (TZBD) voltage for flash-annealed gate stacks is higher than that for spike-annealed gate stacks. This might be due to small differences in the crystalline structure of the high-κ films subjected to annealing with rather different peak temperatures and thermal budgets.

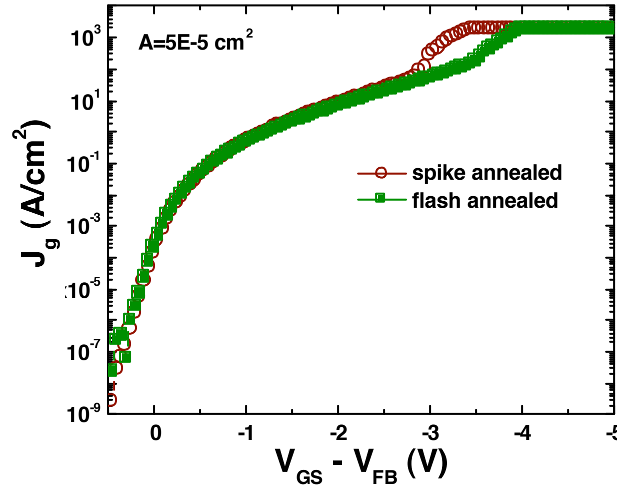


Figure 3.10 Accumulation J_g for both spike and flash annealed gate stacks. Time-zero breakdown (TZBD) voltage is observed to increase for flash annealed gate stacks *vs.* spike annealed gate stacks.

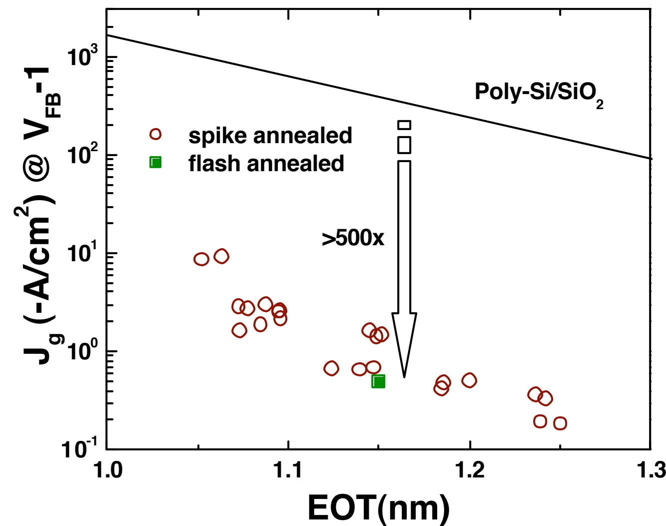


Figure 3.11 $J_{g,accumulation}$ *vs*. EOT for HfSiON/TiN capacitors fabricated with spike or flash annealing. Leakage current density is on par with the historical high-κ trend line. Spike annealed data is taken from [36].

Figure 3.11 shows the relationship between gate current density at $V_{GS} = V_{FB}-1$ V and *EOT*. For reference, the SiO_2/poly-Si model is shown. The flash annealed HfSiON/TiN gate stack data point is in agreement with the trendline for spike annealed gate stacks, and >500× leakage reduction versus SiO_2/poly-Si is achieved at 1.15 nm EOT.

Fast transient charging (FTC) in high-κ dielectrics is one of the major high-κ process integration issues. FTC has been shown to be a major contributor to low mobility, poor reliability and V_{TH} instability [37-41]. The use of HfSiON is shown to reduce the charge trapping effects [25]. In order to study the impact of flash annealing on charge trapping, single pulse I_{DS}-V_{GS} measurements were done on transistors with W/L=10 μm/1 μm. Figure 3.12 shows the normalized drain current (I_{DS}) as a function of time when a trapezoidal pulse is applied to the gate for pulse duration of 200 μs with $t_r = t_f$ =100 ns.

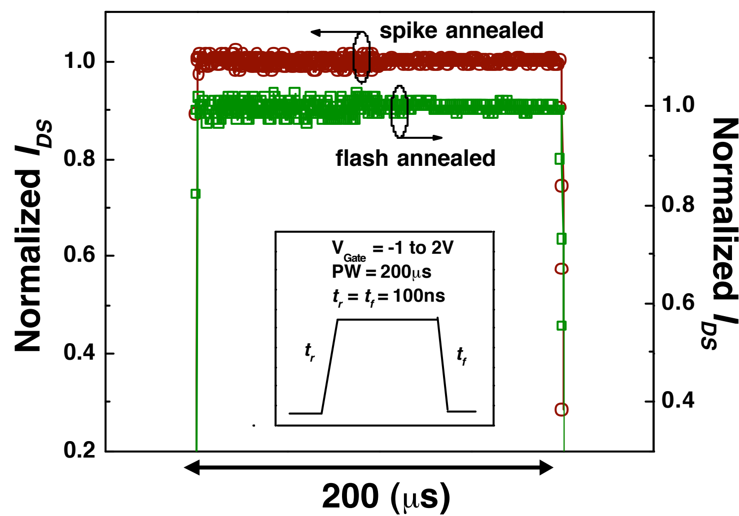


Figure 3.12 Single pulse I_{DS}-V_{GS} measurements show insignificant Δ I_{DS} for both spike and flash annealed gate stacks, indicating negligible fast transient charge trapping. PW, t_r, and t_f correspond to the pulse width, pulse rise, and fall times, respectively.

Monitoring the I_{DS} during the flat pulsewidth portion of the I_{DS}-time plot, minimal I_{DS} drop over the entire pulse duration is observed for the flash annealed sample, indicating negligible charge trapping. This was further confirmed by long-time positive bias temperature instability (PBTI) measurements under different stress voltage biases (V_{GS}-V_{TH} = 1.2 V, 1.4 V). The results indicate negligible PBTI variation with annealing techniques, as shown in Figure 3.13.

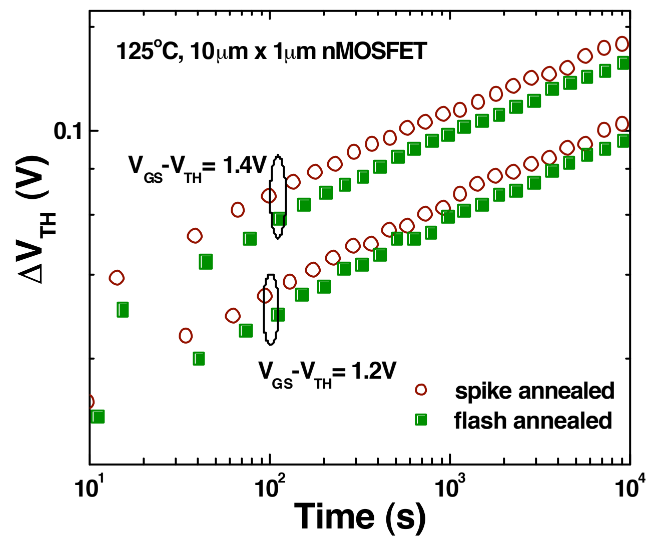


Figure 3.13 Comparison of V_{TH} shift for PBTI stresses at various voltages, for spike *vs.* flash annealed gate stacks.

Therefore, it can be concluded that flash annealing process has minimal effects on the gate stack properties such as *EOT*, V_{FB}, and leakage, and it appears compatible with high-κ/metal-gate stacks. Furthermore, it is evident that bulk charge trapping in high-κ is not an issue with flash annealing.

3.4.3 MOSFET Performance

Linear peak transconductance (G_m) as a function of gate length for spike and flash annealed nMOSFETs (*W*=10 μm) is shown in Figure 3.14. Significant degaradation in

the peak G_m value is observed for flash annealed FETs as the gate length is scaled down to 100 nm; ~1.8× reduction is seen at L_g = 100 nm.

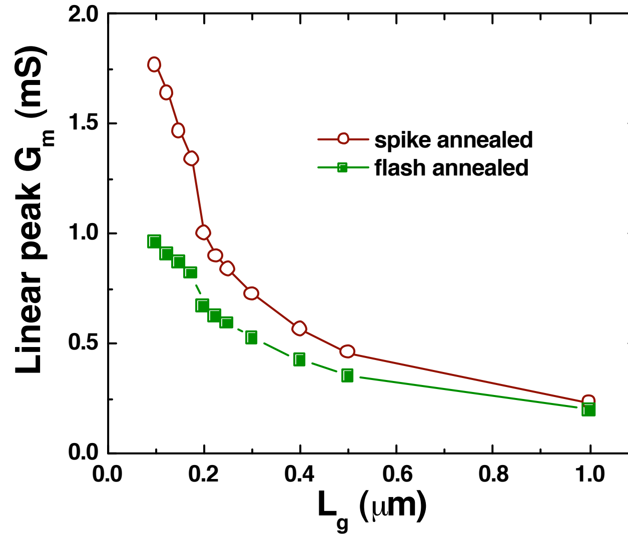


Figure 3.14 Effect of S/D activation anneal technique on G_m, plotted as a function of gate length, L_g. (W =10 μm, V_{DS}=50 mV).

The effective electron mobility was extracted next using drain conductance (g_d) determined at low drain bias and carrier density (N_{inv}) determined by split *C-V* measurement. This extraction was further confirmed by using the NCSU MOB2D model [42]. However, for identical S/D implantation conditions, shallower junction depth and higher dopant activation is achieved for flash annealed MOSFETs as compared to spike annealed MOSFETs. Therefore, different S/D parasitic resistance (R_{sd}) is expected for flash and spike annealed devices. To decouple the impact of R_{sd} on device performance, drain conductance was corrected for R_{sd} ?

$$g_d(R_{sd}) = \frac{g_{d0}}{1 + g_{d0} R_{sd}} \quad (3.2)$$

where g_{d0} is drain conductance without R_{sd} and $g_d(R_{sd})$ is drain conductance with R_{sd}. DC I_{DS}-V_{GS} measurements were done on 10×1 μm^2 nMOSFETs to determine $g_d(R_{sd})$. R_{sd} was extracted by channel-resistance method. g_{d0} was calculated using eqn. (3.2).

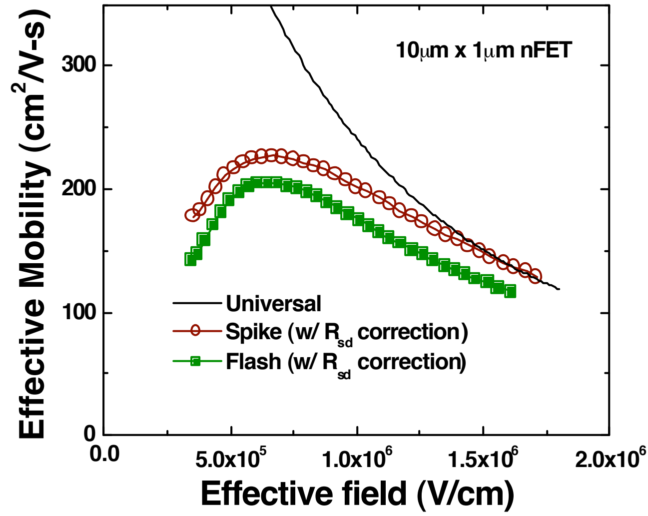


Figure 3.15 Extracted electron mobility *vs*. effective transverse electric field. Mobility is degraded for nMOSFET fabricated with flash annealing *vs*. spike annealing.

The extracted effective electron mobility using g_{d0} and N_{inv} is shown in Figure 3.15. For reference, the universal electron mobility model is also shown. Both the peak and high-field mobility values are degraded for a flash-annealed gate stack. Clearly, the performance of flash annealed MOSFETs is degraded.

Fixed amplitude charge pumping (CP) measurements were done on 10×1 μm^2 nMOSFETs to understand the cause for the mobility degradation. A trapezoidal pulse of amplitude = 1.4 V is applied to the gate with $t_r = t_f$ =10 ns. Interface traps fill from the S/D regions into the channel during t_r and empty into substrate during t_f. Charge pumping current (I_{cp}) is given by

$$I_{cp} = q f A N_{it} \tag{3.3}$$

where f is the frequency, A is the channel area, and N_{it} is the interface trap density. Due to high gate leakage current (EOT is ~1.15 nm), measured charge pumping current was corrected with low frequency I_{cp} (1 kHz) [43]. Figure 3.16 compares the N_{it} values for devices that underwent flash and spike annealing. Significantly higher N_{it} value is observed in the stacks that received flash annealing, which is 3-4× higher versus spike annealed gate stacks.

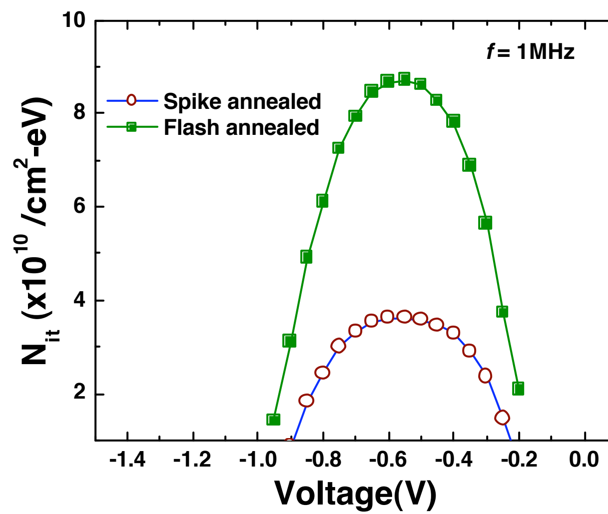


Figure 3.16 Charge pumping (1 MHz) comparison of the two S/D activation annealing methods.

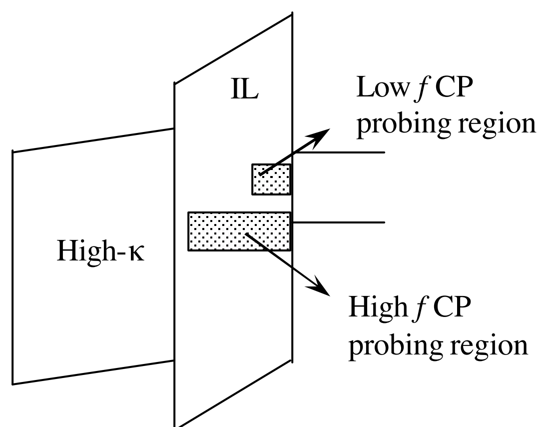


Figure 3.17 Energy band diagram for charge pumping measurements done at different frequencies.

For further investigation, the quality of the interfacial oxide is compared using the charge pumping method at various frequencies, which allows for spatial probing of the interfacial SiO_x layer and high-κ/SiO_x interface as shown in Figure 3.17 [44].

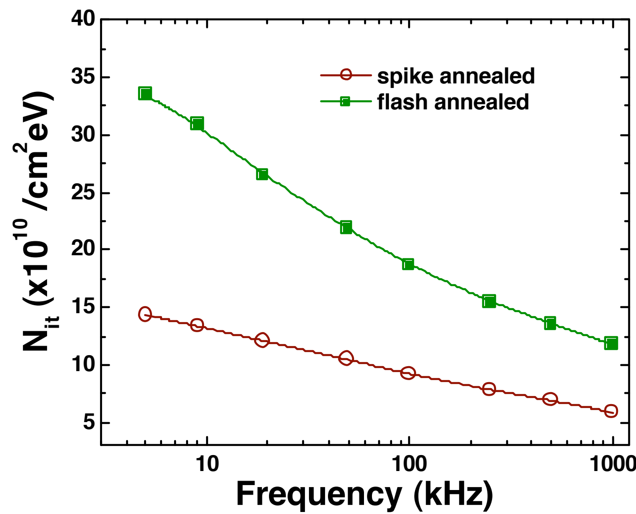


Figure 3.18 Charge pumping measurements of interface trap density (N_{it}) at various frequencies. Higher N_{it} and a significant difference in trap profile are observed for the flash annealed gate stacks as compared to the spike annealed gate stacks.

Figure 3.18 shows measured N_{it} *vs*. frequency for 10×1 μm^2 nMOSFET. The density of traps at high frequencies, which correspond to the trap location near the interface with the substrate, is higher in the flash annealed sample. A higher slope of the N_{it} is also observed for flash annealed gate stacks at lower frequencies, indicating higher trap density towards the HfSiON/SiO_x interface.

Hence, the likely cause of the mobility loss seen in the flash-annealed gate stack is a higher interface state density. This is consistent with the degraded negative bias temperature instability (NBTI) characteristics (Figure 3.19) and higher SS seen for long channel MOSFETs (Figure 3.4).

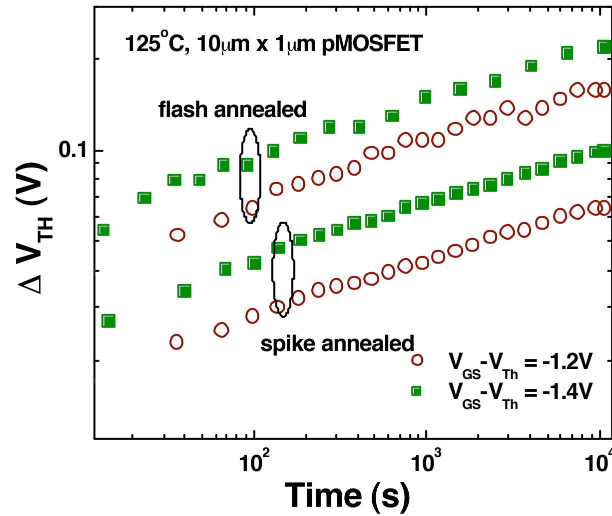


Figure 3.19 NBTI comparison. Degraded characteristics are seen for flash annealed gate stack. This difference is primarily attributed to the degraded Si interface.

To explain the origin of higher N_{it}, wafer-bow measurements were performed before and after the flash-annealing. The radius of curvature measurements are summarized in Table 3.2. A large change in wafer radius of curvature is observed, indicating wafer-level stress after flash annealing. Negative (positive) radius of curvature values correspond to compressive (tensile) stress.

This is not unexpected, because only the device side of the wafer is heated to very high temperature, so that there is a large thermal gradient through the thickness of the wafer, during flash annealing. Similar stress behavior has been reported for a pulsed laser annealing process [31].

Table 3.2 Radius of curvature for the wafers before and after the flash annealing. Negative and positive curvatures correspond to compressive and tensile stress in the wafer, respectively.

Wafer ID	Radius of curvature [m]	Stress
109-02 (before flash anneal)	-471.32	C
109-13 (after flash anneal)	10.32	T
109-14 (after flash anneal)	9.6	T

As has been shown earlier via simulations, CP measurements in the frequency range down to a few kHz may access traps located primarily within the SiO_x interfacial layer and at the high-κ/SiO_x interface [44]. Greater N_{it} change with decreasing CP frequency in the flash annealed devices (seen in Figure 3.18) is thus indicative of structural modification in the SiO_x interfacial layer. At high CP frequencies (1 MHz), when the dielectric is probed in the vicinity of its interface with the substrate, the observed N_{it} increase is probably caused by the thermal-stress-induced defects (broken bonds) at the Si/SiO_x interface. A higher slope of the N_{it} *vs*. frequency trend at lower frequencies corresponding to deeper probing into the gate dielectric indicates a higher density of high-κ induced traps in SiO_x [45]. These defects, which are likely to be oxygen vacancies generated in the SiO_x interfacial layer due to high-κ-driven O out-diffusion [45], can be expected to be created more efficiently during the high-temperature (~1300°C) flash annealing.

3.4.4 Effects of Post-Metallization Anneal

As discussed earlier, flash annealing appears to be compatible with high-κ/metal-gate stacks, but degraded Si interface is seen for flash annealed FETs, resulting in mobility loss. This mobility degradation is a major roadblock for the adoption of flash annealing. In this section, the effects of a post-metallization annealing process on interface properties of fabricated MOSFETs which received flash annealing as S/D activation anneal are presented.

After the conventional forming gas anneal discussed in section 3.4, a high-pressure (10 atm) anneal was performed in pure (100%) deuterium (D_2) ambient for 30 minutes at 400°C [46]. Linear drain current (I_{DS}) and transconductance (G_m) versus gate

voltage (V_{GS}) of 10 μm ×1 μm nMOSFET before and after high-pressure (H.P.) anneal are shown in Figure 3.20. Measurements were taken for V_{DS}=50 mV. After the H.P. anneal, ~7-8% improvement is observed in I_{DS} and G_m.

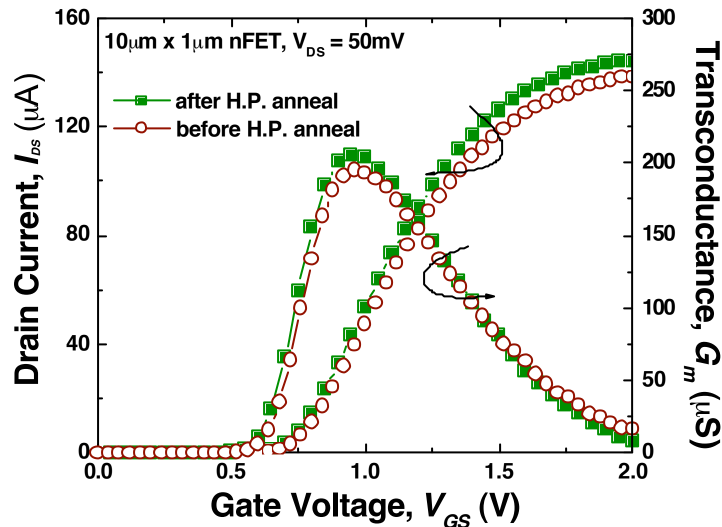


Figure 3.20 Samples annealed in the high-pressure (H.P.) D_2 ambient show improved linear drain current and transconductance (V_{DS}=50 mV).

Capacitance-voltage (*C-V*) measurements were done next on small-area MOS capacitors (A=5×10^{-5} cm^2), and no discernible difference in *EOT* and V_{FB} is seen in *C-V* curves before and after H.P. anneal as shown in Figure 3.21.

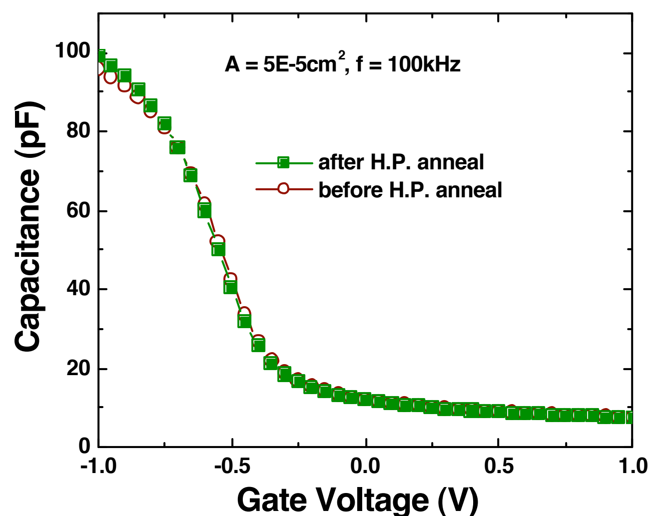


Figure 3.21 Measured high-frequency (f = 100 kHz) *C-V* curves for nMOS capacitors before and after H.P. annealing.

DC I_{DS}-V_{GS} measurements shown in Figure 3.20 were further corrected for series resistance (R_{sd}) and using NCSU MOB2D model, effective electron mobility was extracted. It is found that high-pressure anneal is effective for recovering mobility loss as shown in Figure 3.22. Significant improvement is seen for peak field mobility, which is governed by interface quality. Improved interface is further evidenced by lower subthreshold slope (S.S.) for a 10×1 μm^2 nMOSFET as shown in the Figure 3.20 inset. Thus mobility recovery is attributed to reduction in interface trap density. However, at a higher electric field, mobility gain diminishes, which indicates that H.P. anneal helps to recover mobility loss due to surface roughness scattering, but does not recover it fully. Further optimization of anneal conditions will be needed for further gains.

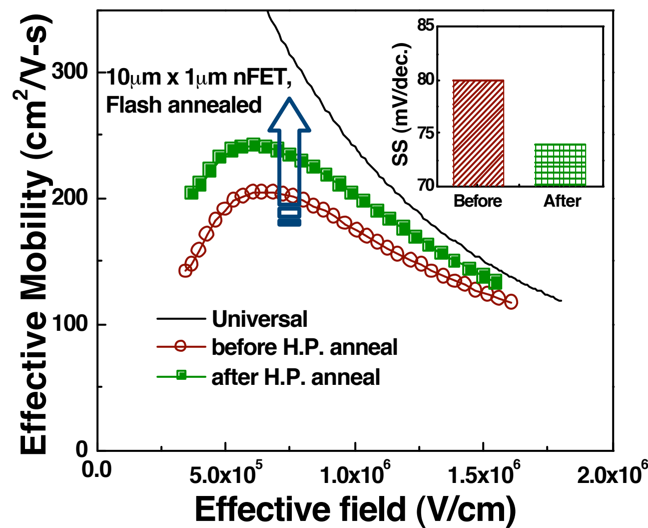


Figure 3.22 Mobility in a flash-annealed device is improved by H.P. anneal. S.S. value is also lowered for 10 μm × 1 μm nMOSFET.

3.4.5 Challenges with Flash Anneal

Flash anneal is a millisecond anneal process. Therefore, its efficacy to annihilate implantation damage is a primary concern because of the shorter anneals time. Figure 3.23 shows the cumulative probability plot for the junction leakage of a *n+/p* diode of

area 10^{-6} cm^2. Comparing the 50% values for flash annealed and spike annealed diode, it is seen that flash annealed junctions leak more as compared to spike annealed junctions by ~10×. The junction leakage is a sensitive function of the location of residual effects with respect to the depletion region, which in turn depends on the S/D extension and halo implant profiles. Therefore, careful optimization of implantation conditions of S/D extension and halo are needed to reduce the junction leakage. In addition, F co-implantation as discussed in section 2.7 can be used to passivate the residual defects and thus reduce the junction leakage.

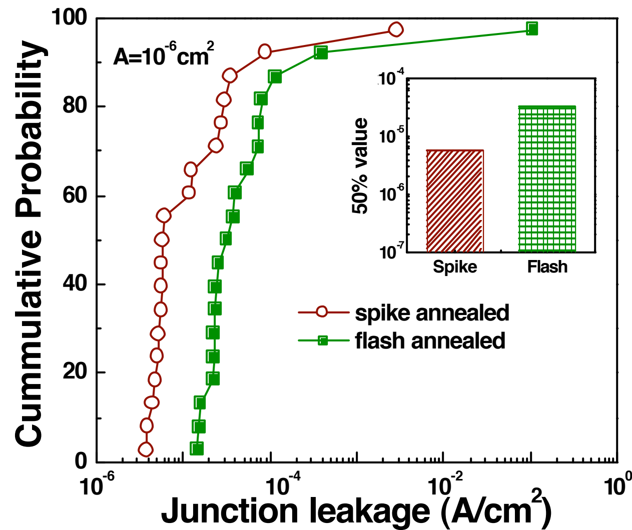


Figure 3.23 Junction leakage of *n+/p* diode after spike or flash anneal.

Figure 3.24 shows the overlap capacitance (C_{ov}) for flash and spike annealed devices. 20% reduction in C_{ov} is seen for flash annealed devices. This reduction can be attributed to shorter anneal time used in flash anneal process. It has been reported that a minimum length of overlap region (L_{ov}) is needed to avoid hot-carrier injection into the ungated region, which leads to reliability problems. Therefore, the small thermal budget of the flash anneal process may become an issue for highly-scaled CMOS devices

because it does not result in enough overlap of the source/drain region under the gate. To counter this problem, an additional low temperature anneal before flash anneal may be needed in conjunction with flash anneal, to provide enough thermal budget to satisfy the L_{ov} requirement. The upper limit of the thermal budget for this anneal will be set by USJ requirements. This additional anneal can also help annihilate implant damage, reducing the higher junction leakage shown in Figure 3.23.

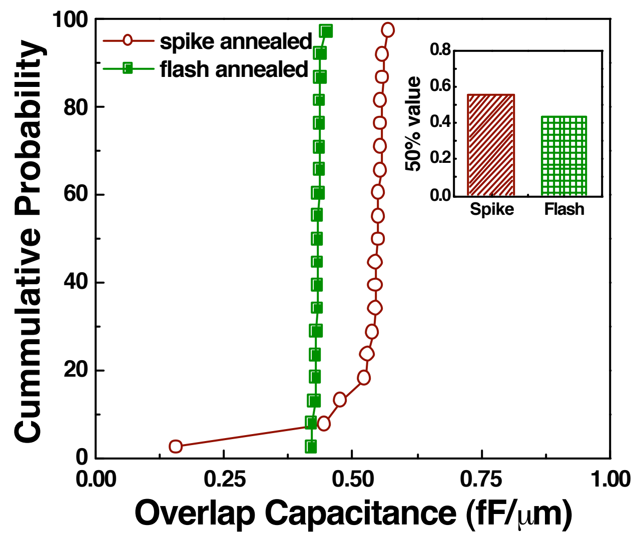


Figure 3.24 Overlap capacitance for flash and spike annealed devices.

3.5 Summary

The process integration of high-κ/metal-gate with flash anneal is presented in this chapter. The flash anneal process is found to be compatible with the high-κ/metal-gate stack in terms of gate stack integrity, and is effective to achieve the benefits of USJs. Generation of O vacancies within the interfacial SiO_x layer due to thermal stress and high-temperature processing results in higher interface-state density and hence degraded mobility. An optimized post-metallization annealing process is shown to be effective for recovering device performance.

3.6 References

[1] S. H. Lo, D. A. Buchanan, Y. Taur, and W. Wang, "Quantum-mechanical Modeling of Electron Tunneling Current from the Inversion Layer of Ultra-thin-oxide nMOSFETs," *IEEE Electron Device Letters*, vol. 18, no. 5, pp. 209 -211, 1997.

[2] S. Song, W. S. Kim, J. S. Lee, T. H. Choe, J. K. Choi, M. S. Kang, U. I. Chung, N. I. Lee, K. Fujihara, H. K. Kang, S. I. Lee, and M. Y. Lee, "Design of sub-100 nm CMOSFETs: Gate Dielectrics and Channel Engineering," *IEEE VLSI Technology Symp. Tech. Digest*, pp. 190-191, 2000.

[3] S. A. Campbell, D. C. Gilmer, X. C. Wang, M. T. Hsieh, H. S. Kim, W. Gladfelter, and J. Yan, "MOSFET Transistors Fabricated with High-permittivity TiO_2 Dielectrics," *IEEE Trans. Electron Devices*, vol. 44, no. 1, pp. 104-109, 1997.

[4] B. H. Lee, Y. Jeon, K. Zawadzki, W. Qi, and J. C. Lee, "Effects of Interfacial Layer Growth on The Electrical Characteristics of Thin Titanium Oxide Films on Silicon," *Applied Physics Letters*, vol. 74, no, 21, pp. 3143-3145, 1999.

[5] C. Hobbs, R. Hedge, B Maiti, H. Tseng, D. Gilmer, P. Toblin, O. Adetutu, F. Huang, D. Weddington, R. Nagabushnam, D. O'Meara, K. Reid, L. La, L. Grove, M. Rossow, "Sub-quarter Micron CMOS Process for TiN-gate MOSFETs with TiO2 Gate Dielectric Formed by Titanium Oxidation," *IEEE VLSI Technology Symp. Tech. Digest*, pp. 133-134, 1999.

[6] A. Chatterjee, R. A. Chapman, K. Joyner, M. Otobe, S. Hattangady, M. Bevan, G. A. Brown, H. Yang, Q. He, D. Rogers, S. J. Fang, R. Kraft, A. L. P. Rotondaro, M. Terry, K. Brennan, S. -W. Aur, J. C. Hu, H. -L. Tsai, P. Jones, G. Wilk, M. Aoki,

M. Rodder, and I. -C. Chen, "CMOS Metal Replacement Gate Transistors Using Tantalum Pentoxide Gate Insulator," *IEEE IEDM Tech. Digest*, pp. 777-780, 1998.

[7] H. F. Luan, S. J. Lee, C. H. Lee, S. C. Song, Y. L. Mao, Y. Senzaki, D. Roberts, and D. L. Kwong, "High Quality Ta_2O_5 Gate Dielectrics with $T_{ox,eq}$<10Å," *IEEE IEDM Tech. Digest*, pp. 141-144, 1999.

[8] Q. Lu, D. Park, A. Kalnitsky, C. Chang, C.-C. Cheng, S. P. Tay, T.-J. King, and C. Hu, "Leakage Current Comparison between Ultra-thin Ta_2O_5 Films and Conventional Gate Dielectrics," *IEEE Electron Device Letters*, vol. 19, no. 9, pp. 341-342, 1998.

[9] S. Guha, E. Cartier, M. A. Gribelyuk, N. A. Bojarczuk, and M. A. Copel, "Atomic beam deposition of lanthanum and yttrium-based oxide thin films for gate dielectrics," *Applied Physics Letters*, vol. 77, no. 17, pp. 2710-2712, 2000.

[10] J. J. Chamber and G. N. Parson, "Yttrium Silicate Formation on Silicon: Effect of Silicon Preoxidation and Nitridation on Interface Reaction Kinetics," *Applied Physics Letters*, vol. 77, no. 15, pp. 2385-2387, 2000.

[11] R. A. McKee, F. J. Walker, and M. F. Chisholm, "Crystalline Oxide on Silicon: The First Five Monolayers, " *Physical Review Letters*, vol. 81, no. 14, pp. 3014-3017, 1998.

[12] J. H. Lee, K. Koh., N. I. Lee, M. H. Cho, Y. K. Kim, J. S. Jeon, K. H. Cho, H. S. Shin, M. H. Kim, K. Fujihara, H. K. Kang, and J. T. Moon, "Effect of Polysilicon Gate on The Flatband Voltage Shift and Mobility Degradation for Al_2O_3 Gate Dielectric," *IEEE IEDM Tech. Digest*, pp. 645-648, 2000.

[13] D. A. Buchanan, E. P. Gusev, E. Cartier, H. Okorn-Schmidt, K. Rim, M. A. Gribelyuk, A. Mocuta, A. Ajmera, M. Copel, S. Guha, N. Bojarczuk, A.Callegari, C. D'Emic, P. Kozlowski, K. Chan, R. J. Fleming, P. C. Jamison, J. Brown, and R. Arndt, "80 nm Poly-silicon Gated n-FETs with Ultra-thin Al_2O_3 Gate Dielectric for ULSI Applications", *IEEE IEDM Tech. Digest*, pp. 223-226, 2000.

[14] E. P. Gusev, M. Copel, E. Cartier, I. J. R. Baumvol, C. Krug, and M. A. Gribelyuk, "High-resolution Depth Profiling in Ultrathin Al_2O_3 Films on Si," *Applied Physics Letters*, vol. 76, no. 2, pp. 176-178, 2000.

[15] M. Balog, M. Schieber, M. Michman, and S. Patai, "Chemical Vapor Deposition and Characterization of HfO_2 Films from Organo-hafnium Compounds," *Thin Solid Films*, vol. 41, no. 3. pp. 247-259, 1977.

[16] B. H. Lee, L. Kang, W.-J. Qi, R. Nieh, Y. Jeon, K. Onishi, and J. C. Lee, "Ultrathin Hafnium Oxide with Low Leakage and Excellent Reliability for Alternative Gate Dielectric Application," *IEEE IEDM Tech. Digest*, pp. 133-136, 1999.

[17] M. Balog, M. Schieber, M. Michman, and S. Patai, "The Chemical Vapor Deposition and Characterization of ZrO_2 Films from Organometallix Compounds," *Thin Solid Films*, vol. 41, no. 2, pp. 109-120, 1977.

[18] M. Copel, M. A. Gribelyuk, and E. Gusev, "Structure and Stability of Ultrathin Zirconium Oxide Layers on Si(001)," *Applied Physics Letters*, vol. 76, no. 4, pp. 436-438, 2000.

[19] W. J. Qi, R. Nieh, B. H. Lee, L. Kang, Y. Jeon, K. Onishi, T. Ngai, S. Banerjee, and J. C. Lee, "MOSCAP and MOSFET Characteristics Using ZrO_2 Gate Dielectric Deposited Directly on Si," *IEEE IEDM Tech. Digest*, pp. 145-148, 1999.

[20] H. Wong and H. Iwai, "On the Scaling Issues and high-κ Replacement of Ultrathin Gate Dielectric for Nanoscale MOS Transistors," *Microelectronic Engineering*, vol 83, no. 10, pp. 1867-1904, 2006.

[21] N. V. Nguyen, M. M. Frank, A. V. Davydov, and D. Chandler-Horowitz, "Sub-Bandgap Defect States in Polycrystalline Hafnium Oxide and Their Suppression by Admixture of Silicon" *Applied Physics Letters*, vol. 87, pp. 192903, 2005.

[22] K. Sekine, S. Inumiya, M. Sato, A. Kaneko, K. Eguchi, and Y. Tsunashima, "Nitrogen Profile Control by Plasma Nitridation Technique for Poly-Si Gate HfSiON CMOSFET with excellent Interface Property and Ultra-low Leakage Current," *IEEE IEDM Tech. Digest*, pp. 103-106, 2003.

[23] C. Hobbs, L. Fonseca, V. Dhandapani, S. Samavedam, B. Taylor, J. Grant, L. Dip, D. Triyoso, R. Hegde, D. Gilmer, R. Garcla, D. Roan, L. Lovejoy, R. Rai, L. Hebert, H. Tseng, B. White, and P. Tobin, "Fermi Level Pinning at the Poly-S/Metal Oxide Interface," *IEEE VLSI Technology Symp. Tech. Digest*, pp. 9-10, 2003.

[24] C. C. Hobbs, L. R. C. Fonseca, A. Knizhnik, V. Dhandapani, S. B. Samavedam, W. J. Taylor, J. M. Grant, L. G. Dip, D. H. Triyoso, R. I. Hegde, D. C. Gilmer, R. Garcia, D. Roan, M. L. Lovejoy, R. S. Rai, E. A. Hebert, H.-H. Tseng, S. G. H. Anderson, B. E. White, and P. J. Tobin, "Fermi-level Pinning at the Polysilicon/Metal-Oxide Interface-Part II," *IEEE Transactions on Electron Devices*, vol. 51, no. 6, pp. 978-984, 2004.

[25] M. A. Quevedo-Lopez, S. A. Krishnan, P. D. Kirsch, H. J. Li, J. H. Sim, C. Huffman, J. J. Peterson, B. H. Lee, G. Pant, B. E. Gnade, M. J. Kim, R. M. Wallace,

D. Guo, H. Bu, and T. P. Ma, "High Performance Gate First HfSiON Dielectric Satisfying 45nm Node Requirements," *IEEE IEDM Tech. Digest*, pp. 425-428, 2005.

[26] K. Mistry, C. Allen, C. Auth, B. Beattie, D. Bergstrom, M. Bost, M. Brazier, M. Buehler, A. Cappellani, R. Chau, C. -H. Choi, G. Ding, K. Fischer, T. Ghani, R. Grover, W. Han, D. Hanken, M. Hattendorf, J. He, J. Hicks, R. Huessner, D. Ingerly, P. Jain, R. James, L. Jong, S. Joshi, C. Kenyon, K. Kuhn, K. Lee, H. Liu, J. Maiz, B. Mclntyre, P. Moon, J. Neirynck, S. Pae, C. Parker, D. Parsons, C. Prasad, L. Pipes, M. Prince, P. Ranade, T. Reynolds, J. Sandford, L. Shifren, J. Sebastian, J. Seiple, D. Simon, S. Sivakumar, P. Smith, C. Thomas, T. Troeger, P. Vandervoorn, S. Williams, and K. Zawadzki, " A 45nm Logic Technology with High-k+Metal Gate Transistors, Strained Silicon, 9 Cu Interconnect Layers, 193nm Dry Patterning, and 100% Pb-free Packaging," *IEEE IEDM Tech. Digest*, pp. 247–250, 2007.

[27] T. Ito, K. Suguro, T. Itani, K. Nishinohm, K. Matsuo, and T. Saito, "Improvement of Threshold Voltage Roll-of by Ultra-Shallow Junction Formed by Flash Lamp Annealing," *IEEE VLSI Technology Symp. Tech. Digest*, pp. 53-54, 2003.

[28] K. Adachi, K. Ohuchi, N. Aoki, H. Tusjii, T. Ito, H. Itokawa, K. Matsuo, K. Suguro, Y. Honguh, N. Tamaoki, K. Ishimaru, and H. Ishiuchi, "Issues and Optimization of Milisecond Anneal Process for 45 nm node and Beyond," *IEEE VLSI Technology Symp. Tech. Digest*, pp. 142-143, 2005.

[29] A. Shima, Y. Wang, D. Upadhyaya, L. Feng, S. Talwar, and A. Hiraia, "Dopant Profile Engineering of CMOS Devices Formed by Non-Melt Laser Spike Annealing," *IEEE VLSI Technology Symp. Tech. Digest*, pp. 144-145, 2005.

[30] A. Pouydebasque, B. Dumont, S. Denorme, F. Wacquant, M. Bidaud, C. Laviron, A. Halimaoui, C. Chaton, J. D. Chapon, P. Gouraud, F. Leverd, H. Bernard, S. Warrick, D. Delille, K. Romanjek, R. Gwoziecki, N. Planes, S. Vadot, I. Pouilloux, F. Arnaud, F. Boeuf, T. Skotnicki, "High Density and High Speed SRAM Bit-Cells and Ring Oscillators due to Laser Annealing for 45 nm Bulk CMOS," *IEEE IEDM Tech. Digest*, pp. 663-666 , 2005.

[31] S. Severi, E. Augendre, S. Thirupapuliyur, K. Ahmed, S. Felch, V. Parihar, F. Nouri, T. Hoffman, T. Noda, B. O' Sullivan, J. Ramos, E. San Andres, L. Pantisano, A. De Keersgieter, R. Schreutelkamp, D. Jennings, S. Mahapatra, V. Moroz, K. De Meyer, P. Absil, M. Jurczak, S. Biesemans, "Optimization of Sub-melt Laser Anneal: Performance and Relaibility," *IEEE IEDM Tech. Digest*, pp. 859-862, 2006.

[32] D. C. Gilmer, J. K. Schaeffer, W. J. Taylor, G. Spencer, D. H. Triyoso, M. Raymond, D. Roan, J. Smith, C. Capasso, R. I. Hegde, and S. B. Samavedam, "LASER Anneal to Enable Ultimate CMOS Scaling with PMOS Band Edge Metal Gate/High-κ Stacks," *ESSDERC Proc.*, pp. 351-354, 2006.

[33] M. A. Quevedo-Lopez, P. D. Kirsch, S. A. Krishnan, H. N. Alshareef, J. Barnett, H. R. Harris, A. Neugroschel, F. S. Aguirre-Tostado, B. E. Gnade, M. J. Kim, R. M. Wallace, and B. H. Lee, "Systematic Gate Stack Optimization to Maximize Mobility with HfSiON EOT Scaling, " *ESSDERC Proc.*, pp. 113-116 , 2006.

[34] P. D. Kirsch, M. A. Quevedo-Lopez, H. -J. Li, Y. Senzaki, J. J. Peterson, S. C. Song, S. A. Krishnan, N. Moumen, J. Barnett, G. Bersuker, P. Y. Hung, B. H. Lee, T. Lafford, Q. Wang, D. Gay, and J. G. Ekerdt, "Nucleation and Growth Study of Atomic Layer Deposited HfO_2 Gate Dielectrics Resulting in Improved Scaling and Electron Mobility," *J. Applied Physics*, vol. 99, no. 2, pp. 023508, 2006.

[35] J. R. Hauser and K. Ahmed, "Characterization of Ultra-Thin Oxides Using Electrical C-V and I-V Measurements," *Int. Conf. Characterization and Metrology for ULSI Technology*, vol. 449, pp. 235-239, 1998.

[36] P. D. Kirsch, M. A. Quevedo-Lopez, S. A. Krishnan, C. Krug, H. AlShareef, C. S. Park, R. Harris, N. Moumen, A. Neugroschel, G. Bersuker, B. H. Lee, J. G. Wang, G. Pant, B. E. Gnade, M. J. Kim, R. M. Wallace, J. S. Jur, D. J. Lichtenwalner, A. I. Kingon, and R. Jammy, "Band Edge n-MOSFETs with High-κ/Metal Gate Stacks Scaled to EOT=0.9nm with Excellent Carrier Mobility and High Temperature Stability," *IEEE IEDM Tech. Digest*, pp. 629-632 , 2006

[37] L. Pantisano, E. Cartier, A. Kerber, R. Degraeve, M. Lorenzini, M. Rosmeulen, G. Groeseneken, H. E. Maes, "Dynamics of Threshold Voltage Instability in Stacked High-κ Dielectrics: Role of The Interfacial Oxide," *IEEE VLSI Technology Symp. Tech. Digest*, pp. 163-164, 2003.

[38] E. Gusev and C. P. D'Emic "Charge Detrapping in HfO_2 High-κ Gate Dielectric Stack, " *Applied Physics Letters*, vol. 83, no. 25, pp. 5223-5225, 2003.

[39] S. Zafar, A. Callegari, E. Gusev, and M. Fischetti, "Charge Trapping Related Threshold Voltage Instabilities in High Permittivity Gate Dielectric Stacks," *J. Applied Physics*, vol. 93, no. 11, pp. 9298-9304, 2003.

[40] A. Kerber, E. Cartier, L. Pantisnao, M. Rosmeulen, R. Degraeve, T. Kauerauf, and G. Groeseneken, H. E. Maes, and U. Schwalke, "Characterization of the V_T instability in SiO_2/HfO_2 gate dielectrics," *IEEE IRPS Proc*., pp. 41-45, 2003.

[41] A. Kerber, E. Cartier, L. A. Ragnarsson, M. Rosmeulen, L. Pantisnao, R. Degraeve, Y. Kim, and G. Groeseneken, "Direct Measurement of The Inversion Charge in MOSFETs: Application to Mobility Extraction in Alterative Gate Dielectrics," *IEEE VLSI Technology Symp. Tech. Digest*, pp. 159-160, 2003.

[42] J. R. Hauser, "Extraction of experimental mobility data for MOS devices," *IEEE Transactions on Electron Devices*, vol. 43, no. 11, pp. 1981-1988, 1996.

[43] P. Masson, J.-L. Autran and J. Brini, "On The Tunneling Component of Charge Pumping Current in Ultrathin Gate Oxide MOSFETs," *IEEE Electron Devices Letters*, vol. 20, no. 2, pp. 92-94, 1999.

[44] D. Hei, C. D. Young, G. A. Brown, P. Y. Hung, A. Diebold, E. M. Vogel, J. B. Bernstein, and G. Bersuker, "Spatial Distributions of Trapping Centers in HfO_2/SiO_2 Gate Stack," *IEEE Transactions on Electron Devices*, vol. 54, no. 6, pp. 1338-1344, 2007.

[45] G. Bersuker, C. S. Park, J. Barnett, P. S. Lysaght, P. D. Kirsch, C. D. Young, R. Choi, B. H. Lee, B. Foran, K. van Benthem, S. J. Pennycook, P. M. Lenanhan, and J. T. Ryan "The Effect of Interfacial Layer Properties on the Performance of Hf-based Gate Stack Devices," *J. Applied Physics*, vol. 100, no. 9, pp. 094108, 2006.

[46] H. Park, M. S. Rahman, M. Chang, B. H. Lee, R. Choi, C. D. Young, and H. Hwang "Improved Interface Quality and Charge-Trapping Characteristics of MOSFETs

with High-κ Gate Dielectric," *IEEE Electron Devices Letters*, vol. 26, no. 10, pp. 725-727, 2005.

Chapter 4

Low Barrier Height Contact Technology

4.1 Introduction

It is well recognized that one of the critical challenges for nanoscale MOSFET source/drain design is to limit the impact of parasitic source/drain series resistance [1]. As the dimensions of a MOSFET are scaled down to improve the performance and cost per function of the integrated circuit, parasitic series resistance accounts for an increasingly large percentage of the total transistor on-state resistance and thus limits transistor (hence circuit) performance. It has been reported that the resistance of metal silicide contacts to the heavily doped source/drain regions comprises the dominant component of parasitic series resistance in sub-100 nm MOSFETs [2]. The contact resistance (R_{co}) is expressed as:

$$R_{co} = \frac{\rho_c}{A_{cont}} \tag{4.1}$$

where ρ_c is the specific contact resistivity (often referred to as the figure-of-merit for contacts) and A_{cont} is the contact area of the silicide/drain (or silicide/source) interface. With continued transistor (and contact area) scaling, R_{co} is becoming a major issue, as is explained next.

Figure 4.1 shows the scaling of the contacted pitch and the gate-to-gate spacing for Intel's logic technology generations from 90 nm to 32 nm [3, 4]. Figure 4.2 highlights the impact of gate pitch scaling on the transistor drive current [4]. Normalized linear drain current (I_{dlin}) and saturation drain current (I_{dsat}) are plotted against gate-to-gate spacing for four technology generations from 90 nm to 32 nm. Beyond the 45 nm node, gate-pitch scaling dramatically reduces the drive current due to increased resistance with shrinking S/D contact area. However, dramatic gains in I_{dlin} and I_{dsat} are expected if R_{co} can be reduced as shown for the 32 nm technology node in Figure 4.2.

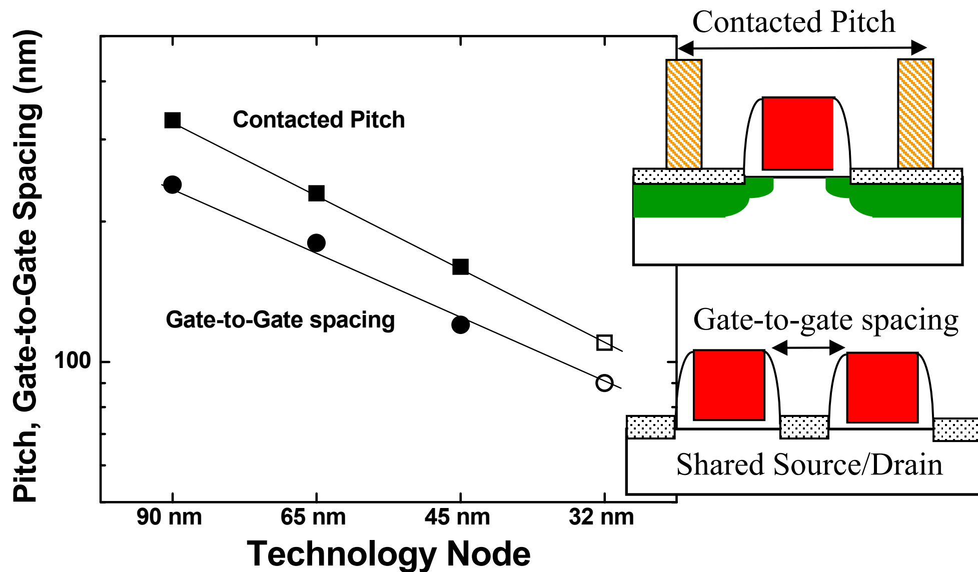


Figure 4.1. The scaling trend of contacted gate pitch and gate-to-gate spacing over the past three technology generations from 90 nm to 45 nm node (shown by closed symbols ■,●). The scaling projections for 32 nm node are shown by open symbols (□,○). The projected value for the pitch is based on 0.7× shrink per generation. Gate-to-gate spacing projection is taken from [4]. Schematic representations of contacted pitch and gate-to-gate spacing are also shown.

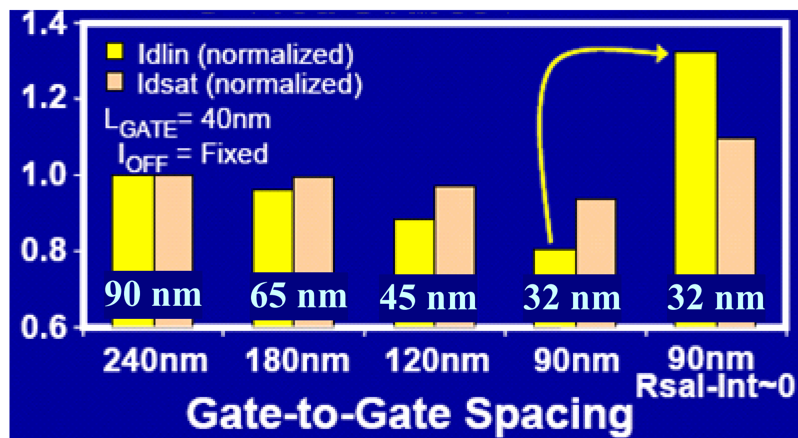


Figure 4.2. The impact of gate-to-gate spacing scaling on transistor drive current in linear and saturation regions [4]. Pitch scaling degrades drive current significantly due to increased resistance. However, with the reduction in silicide-to-S/D contact resistance (or salicide interface resistance labeled as Rsal-Int), performance gains are expected.

Therefore, one key enabler for future transistor performance improvements will be the reduction of ρ_c to limit R_{co}. ITRS projections for ρ_c vs. year of production of bulk high-performance technologies in the near future are shown in Figure 4.3 [1]. ρ_c well

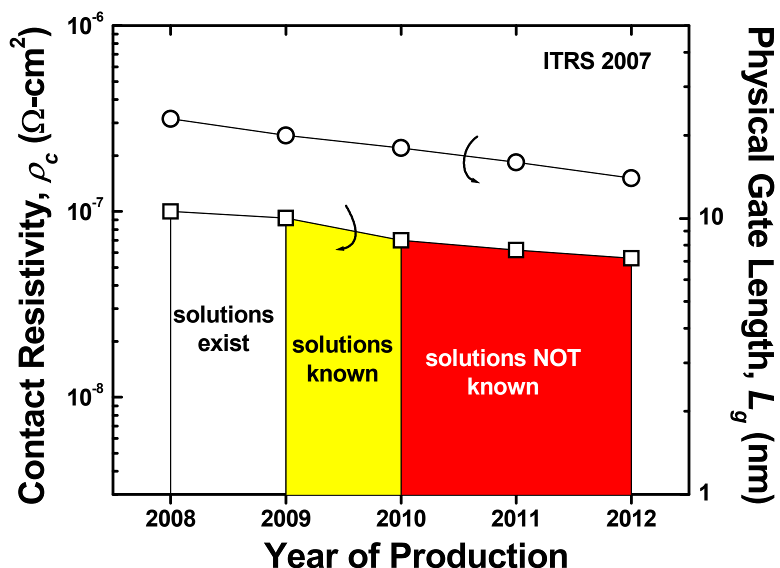


Figure 4.3. Projected scaling trends for contact resistivity (ρ_c) as gate length (L_g) is scaled. The red region indicates that no manufacturable solutions are known to meet these projected values.

below 10^{-7} Ω-cm^2 will be necessary to meet ITRS performance specifications as transistors are scaled to below 20nm gate length (L_g). However, no manufacturable solutions are known to achieve ρ_c projections beyond production year 2010. These projected values overestimate ρ_c requirements because they are calculated assuming that 100% of the allowable MOSFET series resistance is allocated to R_{co}. Clearly, even smaller values of ρ_c will be needed. Thus, solutions for further lowering ρ_c (toward 10^{-8} Ω-cm^2) are needed.

For contacts with high doping, ρ_c is given as:

$$\rho_c \propto \exp\left(\frac{4\pi\sqrt{\varepsilon_s m^*}}{h}\frac{\phi_B}{\sqrt{N}}\right) \tag{4.2}$$

where ϕ_B is the Schottky barrier height of the silicide-Si contact, m^* is the effective mass of the carriers tunneling across the contact, N is the active dopant concentration of the Si region at the interface, h is the Planck constant and ε_s is the dielectric constant of Si. ρ_c depends strongly on ϕ_B and N, and varies exponentially with the factor of $\phi_B/\sqrt{N}$. The calculated values of contact resistivity for n-type silicon, using eqn. (4.2), are plotted in Figure 4.4 as a function of electron barrier height (ϕ_{Bn}) for different values of active dopant concentration. Figure 4.4 indicates that a high active dopant concentration, a low barrier height, or both are necessary to achieve a low value of contact resistivity. For all practical ohmic contacts, higher active dopant concentration has been the primary approach to meet ρ_c requirements for the past technology nodes. The maximum active concentration of dopants is governed by their solid solubility limit in silicon, however. For current CMOS generations, S/D doping is already approaching this limit in silicon. Therefore, sufficiently low ϕ_B (< 0.3 eV) will be needed in order to achieve ρ_c below 10^{-8}

Ω-cm^2. The first potential solution to achieve low ϕ_B is the use of a low workfunction silicide. However, the barrier height of silicide-Si contact is largely determined by the pinning of the Fermi level and not by the workfunction of silicide as discussed next.

In an ideal metal (silicide)-semiconductor junction the Schottky–Mott theory suggests that the electron barrier height, ϕ_{Bn} is determined by the difference between the work function of the metal (silicide), ϕ_M and the electron affinity of the semiconductor, χ:

$$\phi_{Bn} = \phi_M - \chi \tag{4.3}$$

However, it has been observed experimentally that the barrier height is a rather weak function of ϕ_M. This insensitivity can be explained by the presence of unsatisfied covalent bonds (or dangling bonds) in the surface atoms. The dangling bonds give rise to localized energy states at the surface of the semiconductor with energy levels distributed in the bandgap. These surface states affect the barrier height and are characterized by a

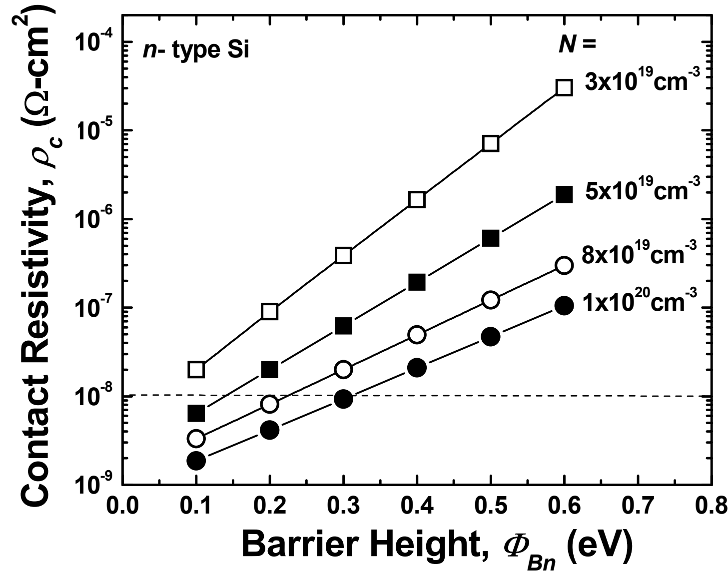


Figure 4.4. The calculated contact resistivity (ρ_c) for *n*-type Si contacts as a function of electron barrier height (ϕ_{Bn}) for different dopant concentration (N). Clearly, significantly low barrier height (below 0.3eV) is required even with very high dopant concentration (~10^{20} cm^{-3}) to achieve $\rho_c < 10^{-8}$ Ω-cm^2.

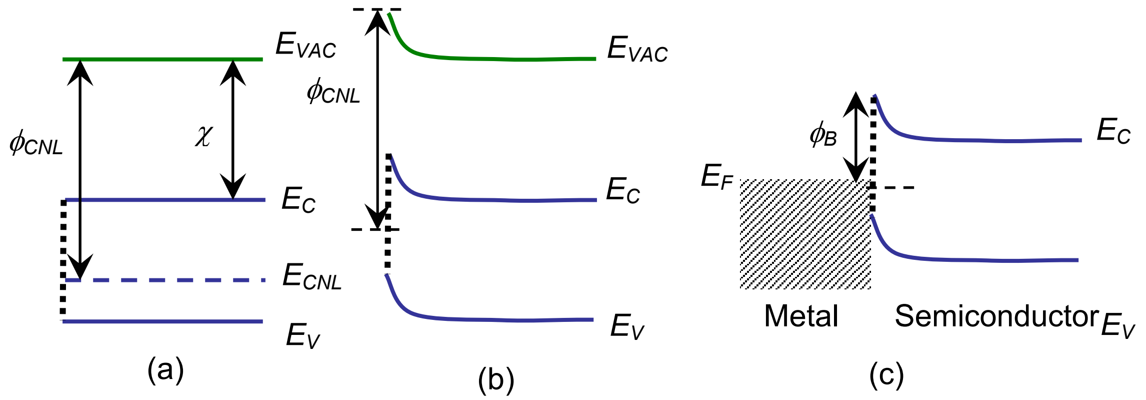


Figure 4.5. The energy band diagrams of *n*-type semiconductor with surface states — (a) flat band at the surface, (b) surface is in equilibrium with the bulk, and (c) semiconductor is in contact with a metal.

charge neutrality level (*CNL*), the energy level below which all surface states must be filled to achieve charge neutrality at the surface. In practice, ϕ_{Bn} of a metal and *n*-type semiconductor contact is expressed as:

$$\phi_{Bn} = S\left(\phi_M - \phi_{CNL}\right) + \left(\phi_{CNL} - \chi\right) \tag{4.4}$$

where ϕ_{CNL} is the energy measured from the *CNL* to the vacuum level as shown in Figure 4.5 and *S* is the slope parameter. The slope parameter

$$S \equiv \left|\frac{\partial \phi_B}{\partial \phi_M}\right| \tag{4.5}$$

provides a measure of the ideality of a metal-semiconductor contact and has a range from 0 to 1. For an ideal metal-semiconductor contact, $S = 1$ and eqn. (4.4) collapses to eqn. (4.3). This situation is called the Schottky limit. When $S = 0$, ϕ_{Bn} becomes constant and ϕ_{Bn} is said to be "pinned" by the surface states near *CNL*. This situation of the strongest pinning is the Bardeen limit. Typically, the *CNL* is located at an energy $\sim 2/3$ of the band gap below the conduction band edge in Si, resulting in higher values of ϕ_{Bn} for metal

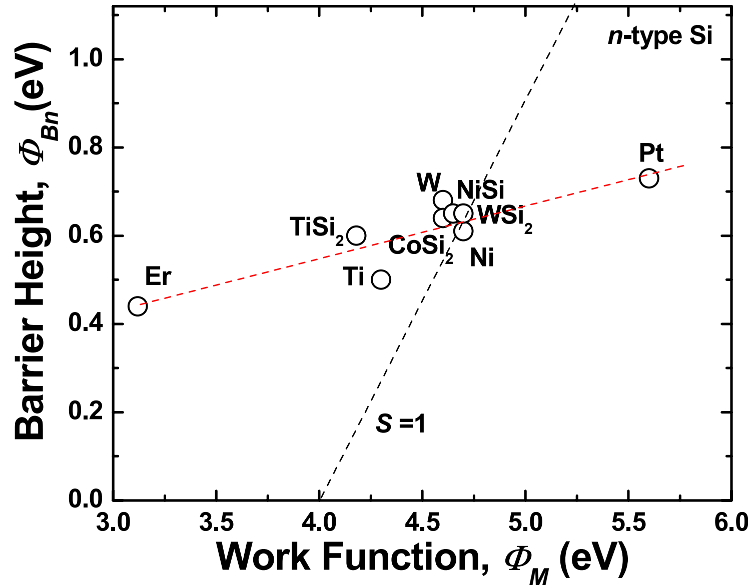


Figure 4.6. Measured electron barrier height as a function of workfunction of various metals and silicides.

(silicide) contact to *n*-type Si. *S* for Si falls in the range 0.1-0.3 due to pinning as shown in Figure 4.6, which shows a plot of the ϕ_{Bn} versus the ϕ_M of the metal or silicide contacting *n*-type Si. Since *S* has values between the Schottky limit and Bardeen limit, barrier height can be controlled to some extent by changing the silicide material. For silicides such as $TiSi_2$, $CoSi_2$ and NiSi that are commonly used to contact heavily doped silicon in modern integrated-circuit devices, however, ϕ_{Bn} is much larger compared to the aforementioned 0.3 eV target, as seen in Figure 4.6. Therefore, the primary challenge for low contact resistance technology is to achieve sufficiently low barrier height.

4.2 Approaches to Reduce Barrier Height

Presently, the understanding of the metal-to-semiconductor barrier height falls short of ideal, and the theory of the gap states due to dangling bonds and the presence of charge neutrality level is largely accepted. However, from the perspective of meeting technological challenges, it is also of great importance to actively search for practical

methods to control the barrier height. Various approaches have been studied in literature to control the barrier height of the silicide to-Si contacts. These approaches are discussed in the following sub-sections.

4.2.1 Bandgap Engineering

The use of bi-axial strain has been proposed to reduce the bandgap and hence the barrier height without disturbance of Fermi-level pinning [5]. The results of theoretical calculations using non-equilibrium Green's function are shown in Figure 4.7. These results indicate that the barrier height for the 2-fold valley (Δ_2) and the light hole band (*LH*) are reduced. A larger change in barrier height for electrons is observed, making this approach particularly attractive for NMOSFETs.

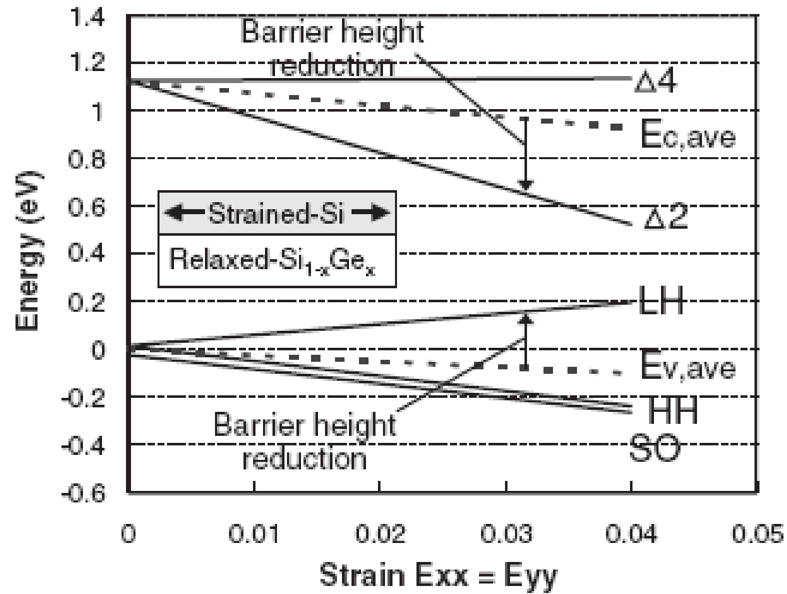


Figure 4.7. The calculated bandgap reduction caused by biaxialstrain [5].

Another possible way to reduce bandgap is to use a smaller bandgap material in the source/drain regions such as $Si_{1-x}Ge_x$. The bandgap of $Si_{1-x}Ge_x$ decreases as the Ge fraction increases. The change in bandgap (ΔE_g) is ~0.1eV @ 25%Ge. This bandgap lowering primarily helps the reduction of hole barrier height (ϕ_{Bn}) as shown in

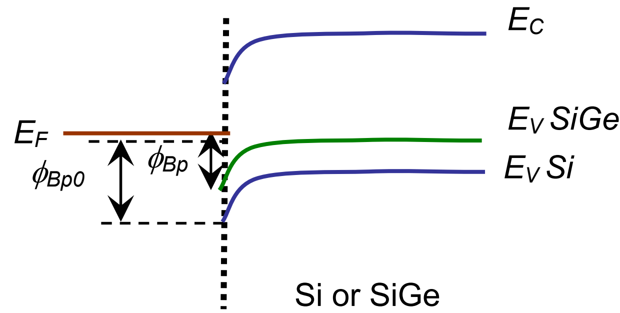


Figure 4.8. The energy band diagrams of *p*-type Si or SiGe in contact with a metal. Due to reduced bandgap of SiGe, barrier height for holes is also reduced.

Figure 4.8, and results in exponential reduction of contact resistance for PMOSFETs. $Si_{1-x}Ge_x$ also offers higher activation of boron relative to Si as shown in Figure 4.9, which further helps to limit the total series resistance. Strained $Si_{1-x}Ge_x$ source/drain technology has now become the industry standard for PMOSFETs in all sub-100 nm high-performance logic nodes because of its multifold benefits.

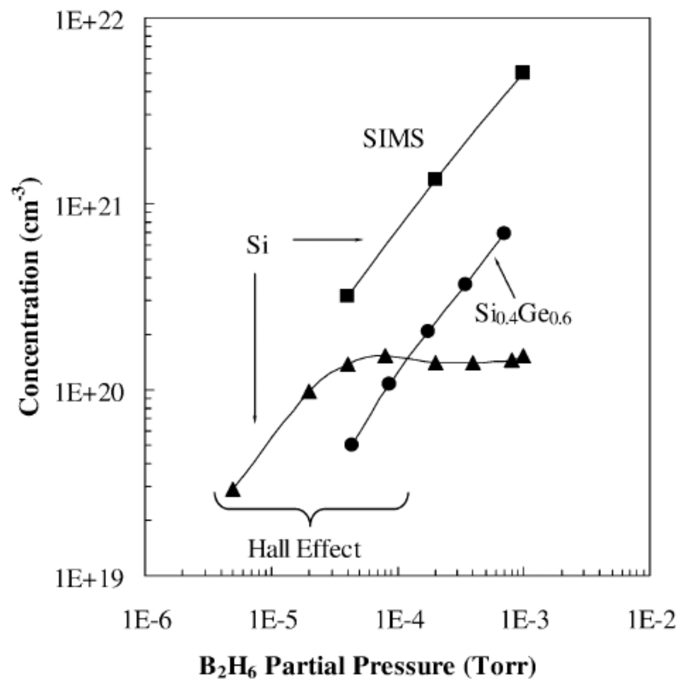


Figure 4.9. The atomic boron concentration in Si measured by SIMS and the active boron concentration in Si and $Si_{0.4}Ge_{0.6}$ obtained by Hall measurements [6].

4.2.2 Dual Silicide Technology

Since the electron and hole barrier heights for each particular silicide add up to the bandgap of Si (1.1eV), it becomes impossible to reduce the barrier for both electrons and holes with one silicide. Reducing the barrier for electrons will increase the barrier for holes and vice versa. Therefore, the silicides used to contact source/drain in future MOSFETs must be separately optimized for NMOS *vs.* PMOS, to maximize CMOS performance. Low-ϕ_M silicide, with a low ϕ_{Bn}, must be used in NMOSFETs and high-ϕ_M silicide, with a low ϕ_{Bp}, must be used in PMOSFETs as shown in Figure 4.10. A dual low-barrier fully-silicided source/drain technology has been reported for thin-body devices using PtSi for PMOS (ϕ_{Bp} = 0.22 eV) and $ErSi_{1.7}$ for NMOS (ϕ_{Bn} = 0.3 eV) [7]. A similar approach can be implemented to achieve low R_{co} technology in a conventional MOSFET structure without the need to fully silicide the source/drain regions.

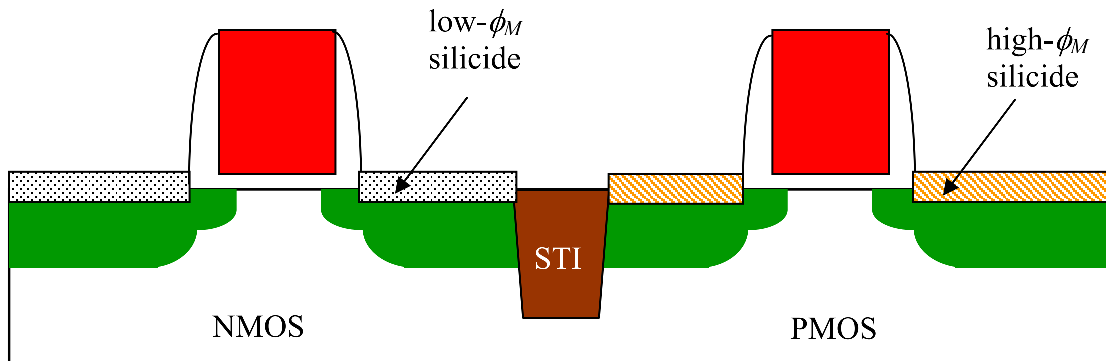


Figure 4.10. Schematic of dual silicide technology employing low-ϕ_M and high-ϕ_M silicides for NMOSFET and PMOSFET, respectively.

4.2.3 Interface Passivation

The other approach is to make use of the dependence of barrier height on the interface structure. Figure 4.11(a) describes the formation of Si(001):(2×1) reconstruction

for minimizing the surface energy. Each atom on a reconstructed Si(001):(2×1) surface has one dangling bond per silicon atom and shares a dimer bond with one neighboring surface atom. As discussed earlier, interface states associated with these dangling bonds lead to Fermi level pinning. By passivating the dangling bonds, S can be improved so that low ϕ_B is achieved in practice with a suitable choice of metal (silicide). A monolayer of selenium (Se) formed by moleular beam epitaxy (MBE) has been shown to effectively passivate the reconstructed Si(001):(2×1) surface by valence-mending as shown in Figure 4.11(b), to enable near-ideal ϕ_{Bn} to be achieved for aluminum and chromium contact to n-type Si [8].

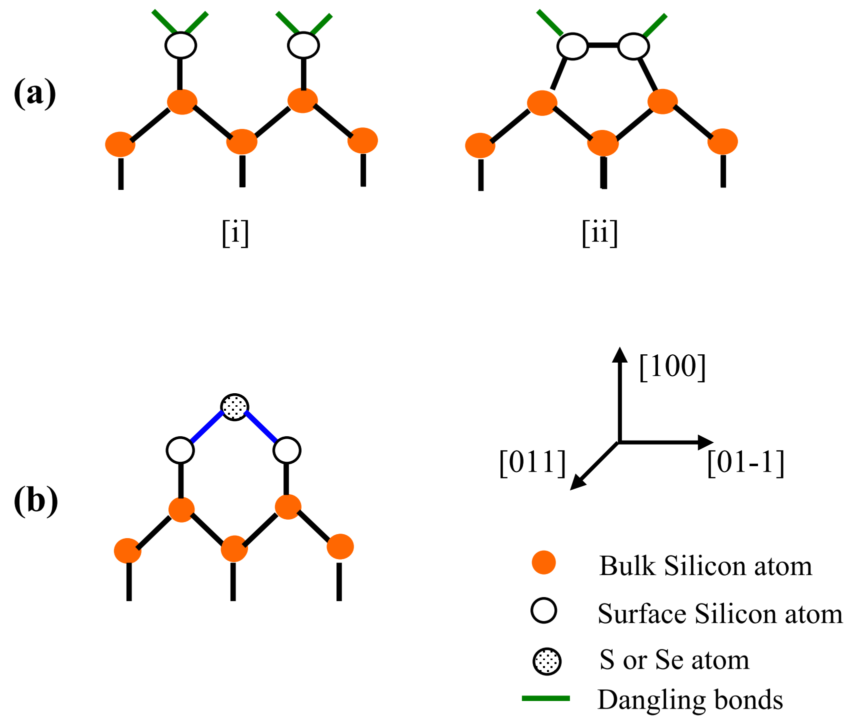


Figure 4.11. (a) Si (001) surfaces [i] ideal bulk terminated: 1×1, [ii] 2×1 reconstruction; (b) Se or S passivated Si (001): 1×1.

ϕ_M can also be de-pinned by inserting an ultra-thin insulator layer between the metal and the semiconductor. The insulator reduces the thermionic barrier by reducing

the gap states density (and depinning the Fermi level) as shown in Figure 4.12. Low ϕ_B on Si was achieved with ~1 monolayer of SiN_x [9].

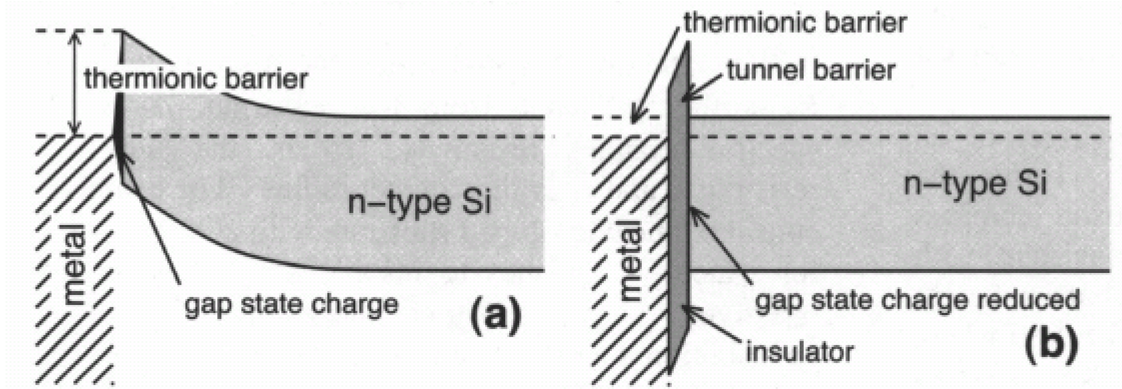


Figure 4.12. Schematic band diagram of metal/*n*-type Si interface (a) without, and (b) with an ultra-thin interfacial insulator [9].

4.2.4 Image-force Induced Barrier Lowering

Another existing scheme to lower ϕ_B is to use image-force induced barrier lowering [10]. Using a heavily doped thin layer (<10 nm) at the metal-semiconductor interface, significant barrier lowering can be achieved with a strong surface electric field as shown in Figure 4.13. Such a thin heavily doped layer can be formed by silicidation-induced dopant segregation [11]. The main advantage of this method is that this layer can be formed at lower temperatures and thus diffusion of dopants from this layer is avoided, resulting in a thin but heavily doped layer at the interface. The dopant segregation technique during silicidation has been demonstrated for fully-silicided $CoSi_2$ bulk SBMOSFETs and fully-silcided NiSi SOI MOSFETs [11, 12].

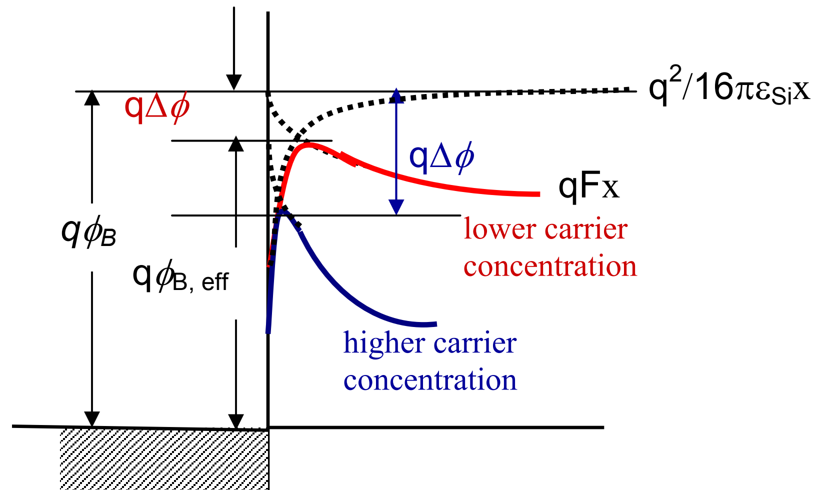


Figure 4.14. Schematic potential profiles depicting the impact of surface carrier concentration on image-force induced barrier lowering. The potential is given by the minimum of field potential (qFx) and image-force potential ($q^2/16\pi\varepsilon_{Si}x$).

For use in fully-silicided MOSFET technology, the doped layer needs to be fully depleted so that the silicide-channel junction retains the Schottky junction properties. However, for low contact resistance technology, no such condition is needed and this technique can be easily implemented in a conventional CMOS flow by the shallow implantation of dopants in the source/drain contact regions (in addition to the standard deep source/drain implant). The silicide thickness must be chosen appropriately to ensure that most of these implanted dopants are segregated to the interface.

4.2.5 Workfunction Engineering

The final way to reduce the barrier height between the silicide and silicon is to engineer the workfunction of the silicide at the interface. This can be achieved by the interfacial segregation of appropriate impurities such as sulfur or selenium as shown in Figure 4.14 [13-15]. However, the mechanism of reported barrier height reduction is not clear, and is believed to be the altered ϕ_M of NiSi and/or the formation of an interface

dipole. Additionally, there are some recent reports of using Ni-alloy silicides [16] and interfacial segregation of metals (e.g. Y, Yb, Pt) for barrier height lowering [17].

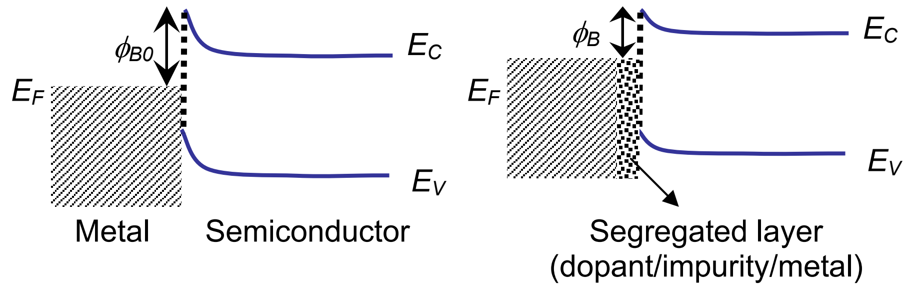


Figure 4.14. The energy band diagrams of metal-semiconductor interfaces without and with a segregated layer.

For future near-term CMOS generations, $Si_{1-x}Ge_x$ S/D technology is expected to provide reduced R_{co} in PMOSFETs with increasingly higher Ge content. However, the challenge of a easily manufacturable low R_{co} technology for NMOSFETs still remains, and requires further investigation. An interface passivation approach such as Se MBE or ultra-thin SiN is attractive, but MBE is a relatively expensive technique, and precise thickness control is needed for SiN_x, to avoid large contact resistance due to tunneling-limited conduction. The integrated process for a dual silicide technology is rather complex and requires a major modification to the conventional CMOS process flow. In comparison, the technique of dopant/impurity/metal segregation to the silicide-silicon interface for barrier height modulation (by image-force barrier lowering or ϕ_M engineering) appears to be the most viable technological option for large-scale integration.

In this work, nitrogen and fluorine are investigated for the passivation of silicide-silicon interface defects. These species are known to passivate grain boundary defects in

polycrystalline silicon TFTs [18-20]. It is possible that a similar effect will be seen for the silicide-silicon interface. The primary advantages of using nitrogen and fluorine are that these species are readily available and are CMOS integration friendly. These species are introduced into silicon prior to silicidation via ion implantation and then segregated to the interface during the silicidation process. Nickel monosilicide (NiSi) is employed in this work because NiSi has emerged as the leading choice for contacting source/drain regions due to its low processing temperature, lower silicon consumption and immunity to the narrow line effect (Table 4.1). Experiments are also carried out using sulfur and selenium implantation for comparison. Note that dopant (e.g. B, As) and metal (e.g. Y, Pt) segregation to the interface are complementary approaches.

Table 4.1. Commonly used self-aligned silicide [21].

Properties	**$TiSi_2$**	**$CoSi_2$**	**NiSi**
Thin film resistivity (μΩ-cm)	13-20	14-20	14-20
Si consumed per nm of metal (nm)	2.3	3.6	1.8
Formation temperature (^oC)	600-700	600-700	400-600
Schottky barrier height on *n*-type Si	0.60	0.64	0.65

4.3 Experimental Details

Schottky diodes were fabricated on Czochralski-grown *n*-type and *p*-type (100) bulk silicon wafers with resistivity 8-24 Ω-cm as shown in Figure 4.15. First, 450 nm of SiO_2 was thermally grown as the field isolation oxide layer. Next, active-area windows for the Schottky diodes were defined, followed by one of the following segregation implant (SI) species: nitrogen, fluorine, sulfur, and selenium. Table 4.2 lists the SI conditions used for various species studied in this experiment. Control samples did not receive this implant. The wafers then underwent a clean prior to deposition of 15 nm

nickel and 10 nm TiN (capping layer). The implant energy and Ni thickness were selected so that most of the silicon containing the SI dose is consumed during silicidation (R_p+3ΔR_p < $T_{Si_consumed}$ ≅ 28 nm). Next, a two-step nickel silicidation process was carried out: a first rapid thermal anneal (RTA) was done at 300°C for 5 minutes; the TiN capping layer and unreacted Ni (if any) was removed using 4:1 H_2SO_4:H_2O_2; a second RTA was done at 450°C for 30 seconds. To achieve good backside contact, the oxide was removed from the wafer backsides, and then aluminum was deposited.

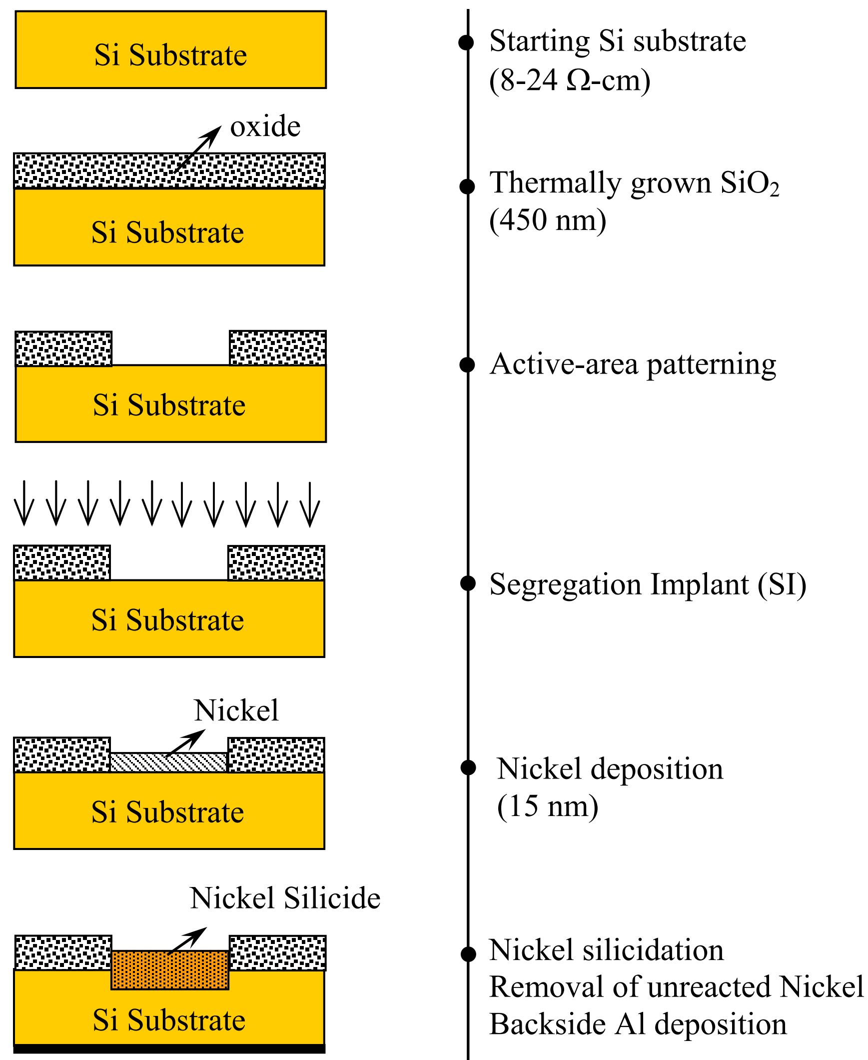


Figure 4.15 Process flow used in this study to fabricate NiSi/lightly doped Si Schottky diodes.

Table 4.2. Implantation energy and dose of various species investigated in this work.

Segregation Implant (SI)	Energy (keV)	Dose range (cm^{-2})	$R_p + 3\Delta R_p$ (nm)
Nitrogen	2	$1\times10^{14} - 3.5\times10^{15}$	24
Fluorine	2.5	$5\times10^{13} - 3.5\times10^{15}$	24
Sulfur	4	$1\times10^{13} - 2\times10^{14}$	21
Selenium	7	$1\times10^{13} - 2\times10^{14}$	22

4.4 Results and Discussion

The Schottky diodes were characterized using current–voltage (*I-V*) measurements at various elevated temperatures. The thermionic currents flowing in a Schottky diode under forward bias and reverse bias are dependent on the absolute temperature, *T*, according to the following equations:

$$\ell n\left(\frac{I_F}{T^2}\right) = \ell n\left(A_0\, A^*\right) - \frac{q}{kT}\left(\phi_B - \frac{V - V_{sub}}{n}\right) \qquad (4.6)$$

$$\ell n\left(\frac{I_R}{T^2}\right) = \ell n\left(A_0\, A^*\right) - \frac{q}{kT}\left(\phi_B\right) \qquad (4.7)$$

where V is the applied bias, I_F is the forward bias current, I_R is the reverse bias current, A_0 is the area of diode, V_{sub} is the voltage drop across bulk silicon substrate due to series resistance, A^* is the Richardson constant, k is the Boltzmann constant, q is the electronic charge, and n is the diode ideality factor. The reverse bias current, given by eqn. (4.7), was used to extract barrier height in order to avoid any error due to series resistance and diode ideality factor. An example of an Arrhenius plot used to extract ϕ_B is shown in Figure 4.16.

Various physical techniques such as X-ray diffraction (XRD), secondary ion mass spectroscopy (SIMS), transmission electron microscopy (TEM), scanning electron

microscopy (SEM) and four-point probe were used to assess the impact of SI process on bulk properties of NiSi.

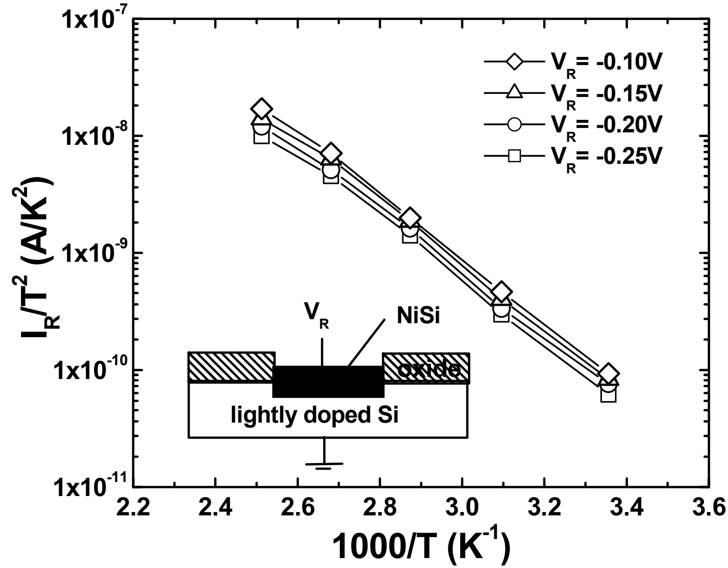


Figure 4.16 NiSi/*n*-type Si diode current *vs.* temperature measurements for extraction of Schottky barrier height, for various reverse bias voltages.

4.4.1 Effects of Nitrogen Implantation

Figure 4.17 shows the nitrogen dose dependence of the Schottky barrier heights ϕ_{Bn} and ϕ_{Bp} for contacts to *n*-type silicon and *p*-type silicon, respectively. It can be seen that ϕ_{Bn} decreases substantially and steadily with an increasing nitrogen dose, from 0.67 eV for the control (unimplanted) sample down to 0.3 eV for the 3.5×10^{15} cm^{-2} implanted sample. In contrast, ϕ_{Bp} increases with an increasing nitrogen dose up to 1×10^{15} cm^{-2}, and saturates at ~0.57 eV for higher doses. Note that for implant doses $\leq 1\times10^{15}$ cm^{-2}, $\phi_{Bn}+\phi_{Bp}$ = 1.1eV, but for higher doses $\phi_{Bn}+\phi_{Bp}$ < 1.1eV, indicating that the silicon band-gap energy is reduced in the samples implanted with doses greater than 1×10^{15} cm^{-2}.

To investigate the impact of nitrogen implantation on the physical structure of silicide/silicon interface, transmission electron microscopy (TEM) analyses were

performed. Figure4.18 (a) shows a cross-sectional high-resolution TEM (HRTEM) image of the sample that received the nitrogen implant. No interface layer is seen based on the HRTEM image. Electron energy loss spectroscopy (EELS) scan acquired across the interface is shown in Figure 4.18 (b); no noticeable nitrogen signal is observed.

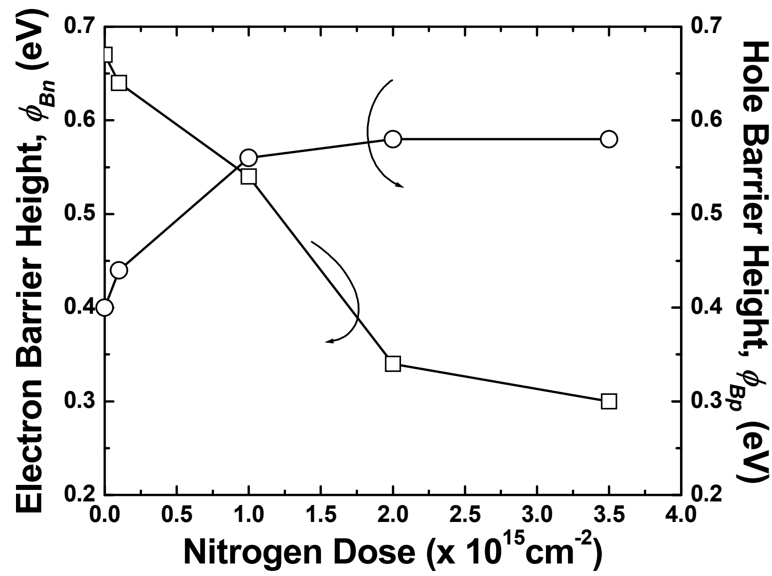


Figure 4.17 Extracted electron and hole Schottky barrier heights as a function of implanted nitrogen dose.

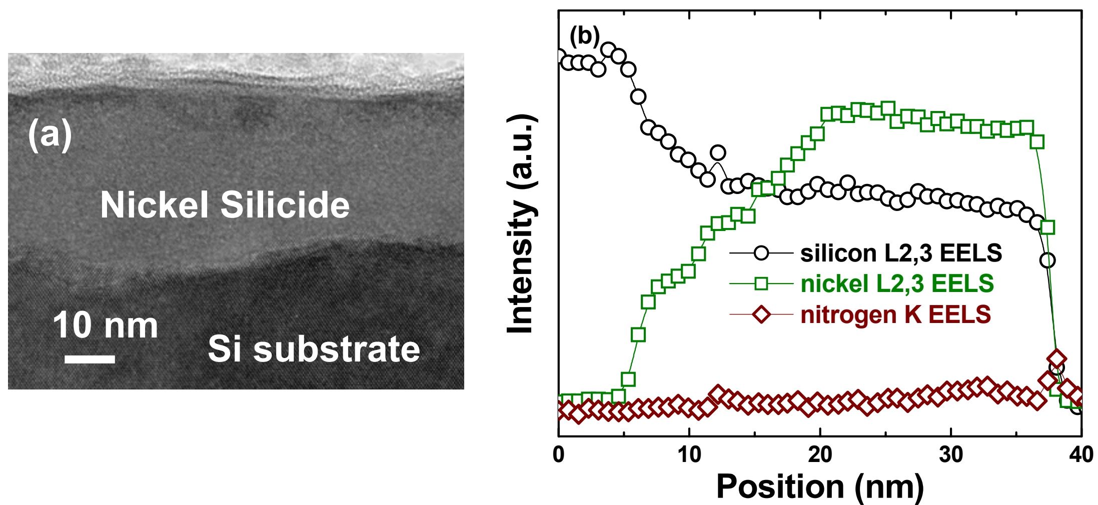


Figure 4.18 (a) Cross-sectional high resolution TEM (HRTEM) micrograph of the sample that received the nitrogen implant. No clear evidence of an interfacial layer is observed, (b) No nitrogen signal is observed in the electron energy loss spectroscopy (EELS) scan at the interface between the Si substrate and the NiSi layer.

Therefore, it can be concluded that nitrogen implantation does not form any interfacial layer (e.g., SiN_x). Hence, the possibility of barrier height modification due to the presence of interfacial SiN_x causing Fermi-level depinning is ruled out.

To elucidate the mechanism of change in Schottky barrier height of N^+ implanted NiSi/Si contacts, glancing-incidence X-ray diffraction (GIXRD), high-resolution X-ray diffraction (HRXRD), and secondary ion mass spectroscopy (SIMS) analyses were carried out. GIXRD 2-theta scans and phi scans provide information on the phase and film texture of silicide, respectively. HRXRD scans characterize the strain in the underlying silicon substrate. SIMS analyses were performed to obtain the nitrogen depth profile.

The GIXRD 2-theta scans shown in Figure 4.19 exhibit peaks at the expected locations for the nickel monosilicide standard, corresponding to the NiSi (011), (111), (112), (211), (103) and (013) planes [22]. Note that there are two strong unassigned peaks at 57.5^o and 83° for the samples which received an implanted nitrogen dose greater than 1×10^{15} cm^{-2}. It is observed that the assigned peaks are shifted slightly to the higher angle as the N^+ dose is increased. The change in 2θ reflects change in *d*-spacing suggesting strain in silicide films. However, GIXRD measures the *d*-spacing of oblique planes and not those parallel to the surface. If it varies, it could be due to either or both (a) mechanical stress or (b) solid solution alloying. Thus, the *d*-spacing change on its own as a measure of stress-induced strain is not conclusive in a system in which there is known to be alloying. Therefore, HRXRD scans were done to further confirm the presence of strain in the samples. The Si (004) rocking curves (RCs) of the HRXRD spectra are shown in Figure 4.20. For implant doses $< 3.5\times10^{15}$ cm^{-2}, there is no major change in

substrate signal. For the sample with the highest dose of nitrogen, RC has a distinct broad peak to the right indicating that a thin layer of silicon is indeed strained compared to the rest of the bulk substrate. Fitting done using the BEDE RADS simulation program estimates the strained layer thickness to be ~162-184Å.

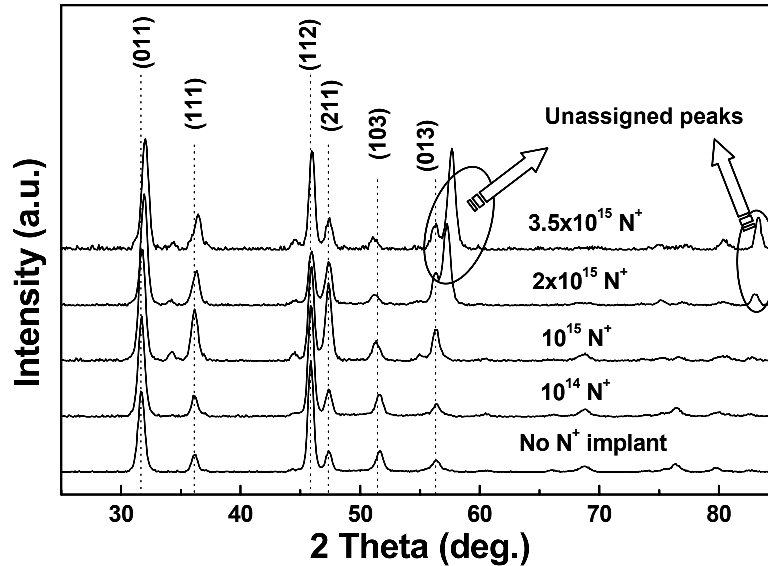


Figure 4.19 GIXRD spectra of nickel silicide films for various nitrogen doses. In addition to NiSi peaks, two strong unassigned peaks located at 57.5° and 83° are also identified for the samples with N^+ dose > 1×10^{15} cm^{-2}.

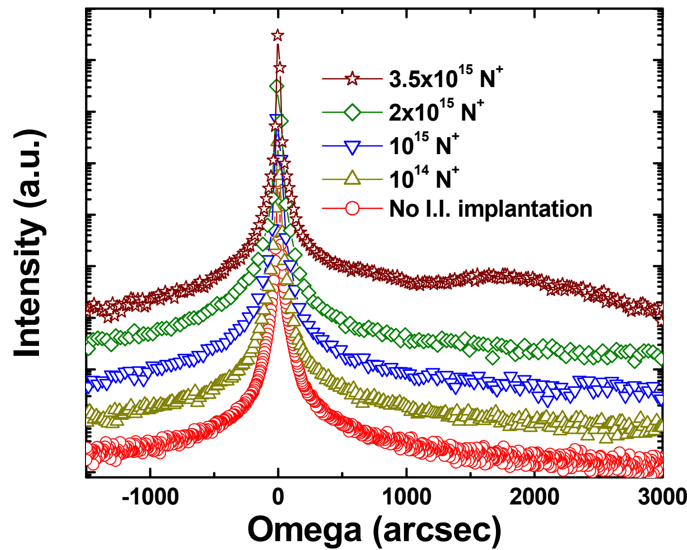


Figure 4.20 HRXRD scans around the Si (004) reflection for the samples with increasing nitrogen dose. For the sample with 3.5×10^{15} cm^{-2} dose, a broad peak is observed to the right.

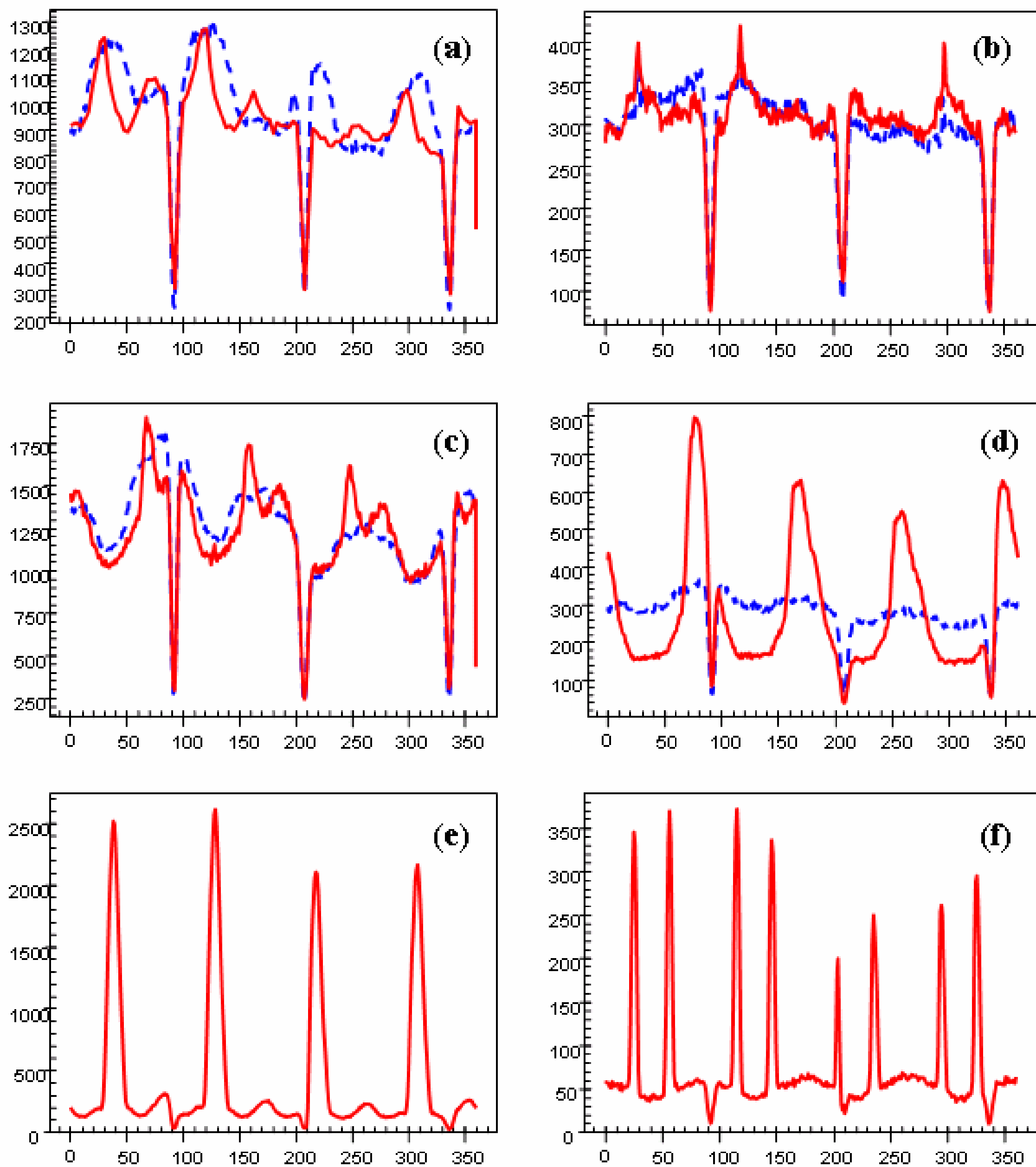


Figure 4.21 GIXRD phi scans for 2θ = (a) 31.6° [NiSi (011)], (b) 36.1° [NiSi (111)], (c) 45.8° [NiSi (112)], (d) 47.3° [NiSi (211)], (e) 57.5° [unassigned] and (f) 83° [unassigned]. Diffraction peaks at 57.5° and 83° are seen only for implanted sample.
X-axis: Phi (deg.); **Y-axis:** Intensity (cps). Blue dotted line represents the sample that received no nitrogen implant (control). Solid red line is for the sample with N^+ dose= 3.5×10^{15} cm^{-2}.

Phi scans using the GIXRD configuration were done next to characterize the texture of nickel silicide films. Figure 4.21 shows the phi scan recorded for the samples without nitrogen implant (control) and with nitrogen implant (dose = 3.5×10^{15} cm^{-2}). Figure 4.21 (a) is the phi scan done for NiSi (011). The difference in the curves from both samples is very small. The sharp spikes present are likely to be due to parasitic motions of the stage, i.e., wobble. Similarly, the data recorded for NiSi (111) is qualitatively the same for both samples as shown in Figure 4.21 (b). However, some differences between the two curves can be seen for NiSi (112) as shown in Figure 4.21 (c). The data from the implanted sample is developing some peaks in the phi scan, which is indicative of preferred orientation. The difference between the two curves becomes very pronounced for NiSi (211). Figure 4.21 (d) shows that the data from the implanted sample contains well developed peaks that are clearly much larger in amplitude than would be expected due to stage wobble. Figures 4.21 (e) & (f) are the phi scans recorded for the peaks at 57.5^o and 83°. Here, very pronounced peaks in the phi scan are seen, which are indicative of some preferred orientation of the NiSi grains in the film. Such strong, well-developed peaks would not be expected from randomly oriented grains (and a bit of stage wobble). Therefore, phi scans for NiSi (211) planes and peaks at 57.5^o and 83° indicate axiotaxy texture in the silicide for the sample with 3.5×10^{15} cm^{-2} implanted dose. Similar indication of axiotaxy texture is also seen for the sample with 2×10^{15} cm^{-2} implanted dose.

The nitrogen depth profiles obtained via SIMS are shown in Figure 4.22. ^{30}Si, $^{28}Si+N$, and ^{64}Ni were monitored as negative ions under 500 eV Cs^+ bombardment at 60°. In order to avoid any artifacts due to knock-on effects, the analysis was done from the

backside. The backside of the sample was first polished until <150 nm of the Si remained. The N concentrations were calculated using a relative sensitivity factor determined from a standard sample. The small peak at the NiSi-Si interface indicates that some of the nitrogen segregated to the interface during the silicidation process. Most of the implanted nitrogen is confined within the NiSi layer, however.

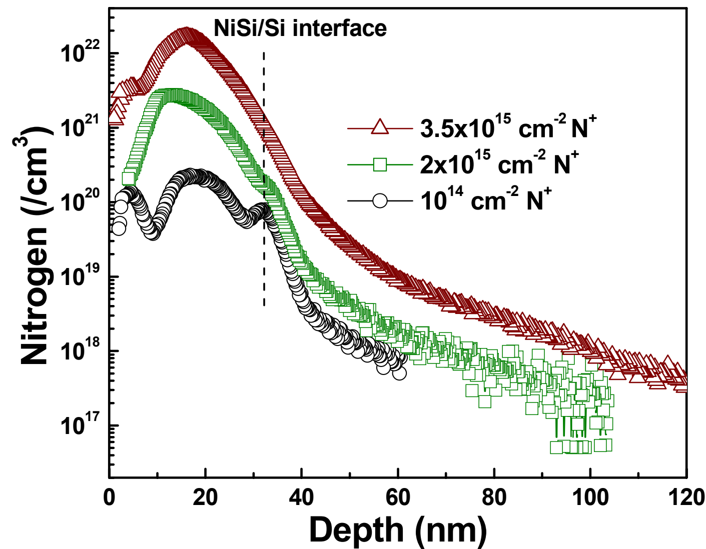


Figure 4.22 Nitrogen depth profiles obtained from backside SIMS analyses. Some nitrogen is seen to be segregated at the NiSi-Si interface, but most of the implanted dose remains confined within the silicide.

With the help of the aforementioned physical analyses, the modification in Schottky barrier height can be explained by two mechanisms: interface passivation at low nitrogen doses, and a combination of strain-induced band-gap narrowing and presence of axiotaxy texture at high nitrogen doses. For doses $\leq 1\times10^{15}$ cm^{-2}, the primary mechanism is the passivation of dangling bonds by nitrogen at the NiSi-Si interface to de-pin the NiSi Fermi level, similar to the effect of Se passivation [8]. For high doses ($>1\times10^{15}$ cm^{-2}), the nitrogen within the NiSi layer induces significant strain in the Si, reducing the

energy band gap. Hence, ϕ_{Bn} is further lowered to 0.3 eV. In the HRXRD scans reported earlier, clear indication of strain is seen for the sample with 3.5×10^{15} cm^{-2} N^+ dose but not for the sample with 2×10^{15} cm^{-2} N^+ dose. It may be possible that the 2×10^{15} cm^{-2} implanted sample is also strained as indicated by bandgap reduction in electrical data, but perhaps the strain is too small to be detected by HRXRD. Another interesting observation is the onset of axiotaxy texture at the implant dose of 2×10^{15} cm^{-2}. For high implant doses ($>1\times10^{15}$ cm^{-2}), ϕ_{Bn} lowering due to the change in NiSi Fermi level is another possible mechanism that may be taking place because of the change in texture.

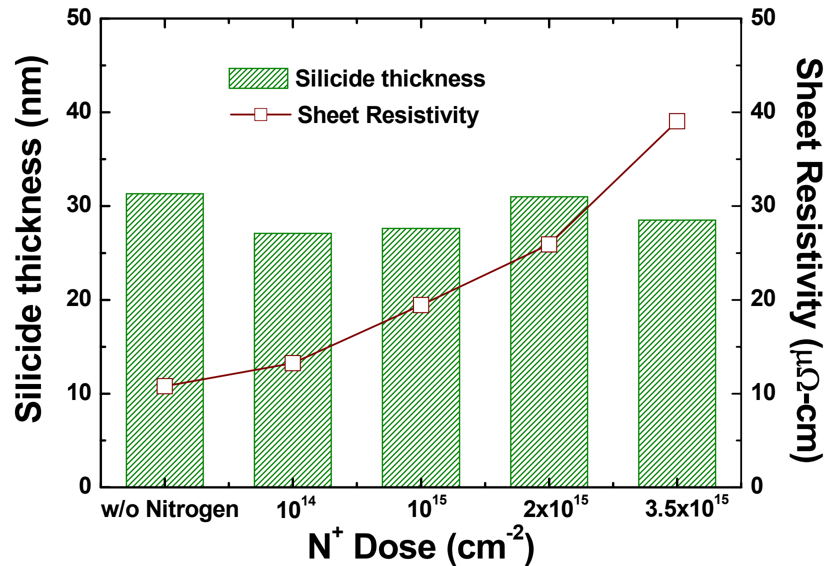


Figure 4.23 Silicide thickness and sheet resistivity *vs.* implanted nitrogen dose.

To investigate the impact of nitrogen incorporation on the NiSi layer, cross-sectional scanning electron microscopy (SEM) analyses and four-point probe measurements were done to determine the silicide thickness and sheet resistance, respectively. Figure 4.23 shows the nitrogen dose dependence of the silicide thickness and the silicide resistivity (ρ_s = sheet resistance × film thickness). There is no significant

change in the silicide thickness as the nitrogen dose is increased. For implant doses ≤ 1×10^{15} cm^{-2}, ρ_s is <20 μΩ-cm, consistent with resistivities reported for NiSi [23, 24]. Unfortunately, ρ_s degrades (increases) for higher implant doses, indicating axiotaxy texture as a possible cause of this degradation.

In summary, it is possible to dramatically reduce the Schottky barrier height for NiSi contact to n-type Si via nitrogen implantation prior to silicidation. A tradeoff between lower ϕ_{Bn} and increased silicide sheet resistivity is seen for high nitrogen doses (>1×10^{15} cm^{-2}), however, so that care must be taken to optimize the implant dose in order to minimize the total parasitic MOSFET source/drain resistance.

4.4.2 Effects of Fluorine Implantation

Figure 4.24 summarizes the extracted electron barrier height for various implanted doses of fluorine. Very small change in ϕ_{Bn} (50-70 meV) is seen even with a higher F^+ implantation dose. The backside SIMS profiles obtained for two different F^+ doses are shown in Figure 4.25 and fluorine is observed to be piled up at the NiSi-Si interface.

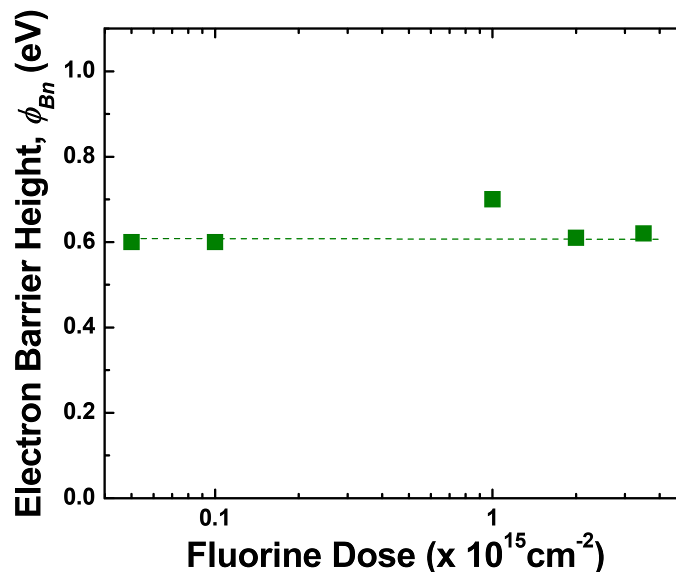


Figure 4.24 Extracted electron barrier heights as a function of implanted fluorine dose.

Figure 4.26 shows the GIXRD 2 theta spectra of the F^+ implanted samples and the diffraction peak positions are in agreement with the NiSi standard. Unlike nitrogen, fluorine segregates well as seen in Figure 4.25 but the ϕ_{Bn} reduction achieved with fluorine implantation is very small compared to nitrogen is not very effective as a passivation specie. This poor efficacy of fluorine passivation is attributed to higher bonding enthalpy of Si-F (576 kJ/mol) compared to that of Si-N (437 kJ/mol).

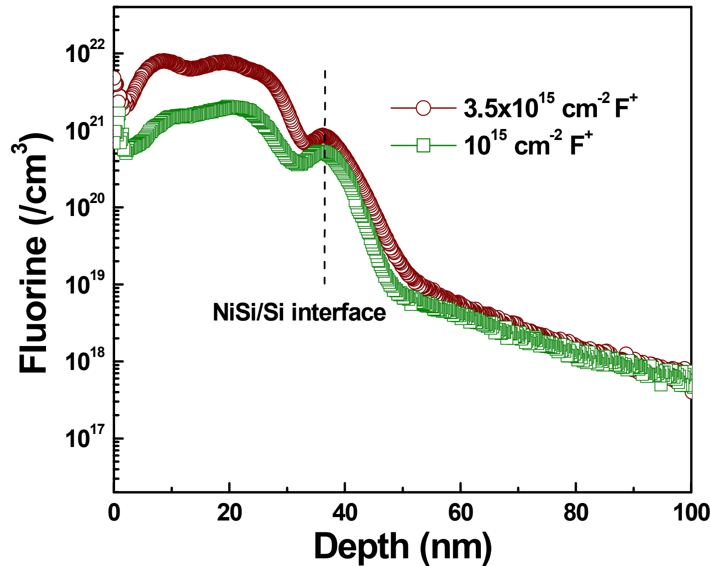


Figure 4.25 Backside SIMS analyses of the NiSi/Si diode samples implanted with fluorine.

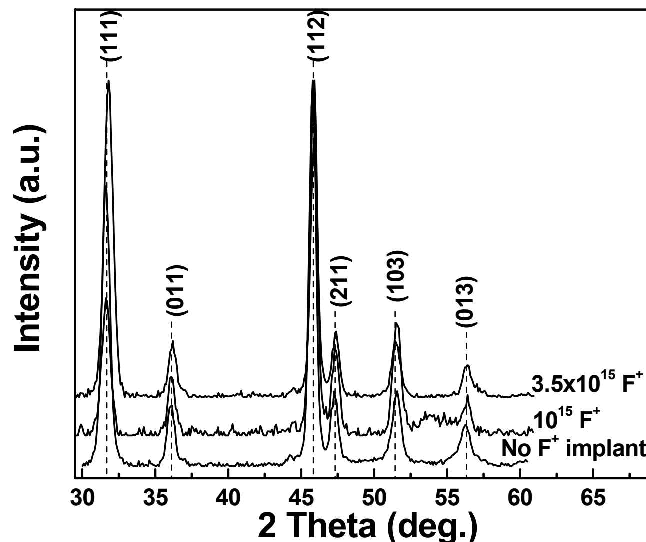


Figure 4.26 GIXRD spectra of nickel silicide films for 10^{15} and 3.5×10^{15} cm^{-2} implanted fluorine doses. GIXRD scan of unimplanted sample (control) is also shown for reference.

4.4.3 Effects of Selenium and Sulfur Implantation

It has been reported that with increasing Se and S segregation at the interface, ϕ_{Bn} below 0.2 eV can be achieved [13, 14]. For the studied conditions in this set of experiments, similar behavior is also observed with increasing implanted Se^+ and S^+ dose, and ϕ_{Bn} is reduced to 0.47 eV for the sample with Se^+ 5×10^{13} cm^{-2} dose (ϕ_{Bn} reduces to 0.58 eV for S^+ 5×10^{13} cm^{-2} dose). However, as the dose is increased to 2×10^{14} cm^{-2} for both Se or S, the slope of the Arrhenius plot becomes positive as shown in Figure 4.27, as opposed to being negative as seen in Figure 4.16.

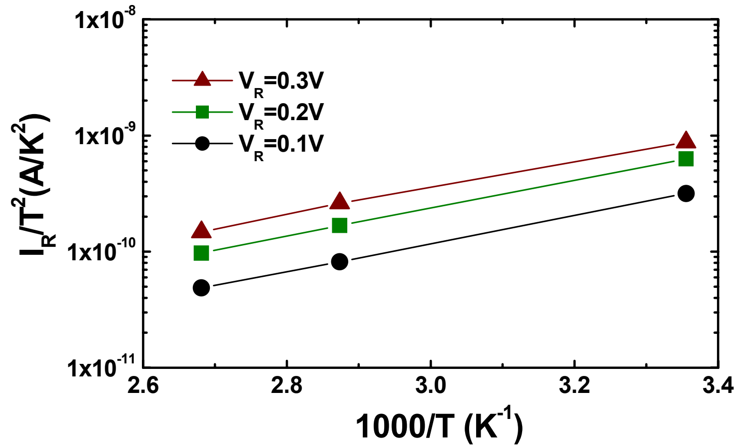


Figure 4.27 The Arrhenius plot for the barrier height determination showing a positive slope for the sample with 2×10^{14} cm^{-2} selenium dose.

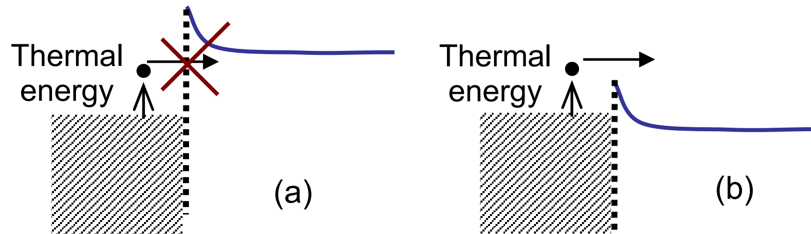


Figure 4.28 The energy band diagrams of a metal-semiconductor junction: (a) thermal energy of carriers is not enough and they see barrier impeding their flow, and (b) carriers have sufficient thermal energy and see no barrier.

This anomaly can be explained by Figure 4.28. The positive slope of the Arrhenius plot suggests that carrier see no barrier impeding their flow. This is possible if the barrier height is very small and carriers have sufficient thermal energy even at room temperature to overcome this barrier as shown in Figure 4.28 (b). However, for a larger barrier height, carriers need to overcome the barrier and negative slope will be seen in the Arrhenius plot (Figure 4.28(a)).

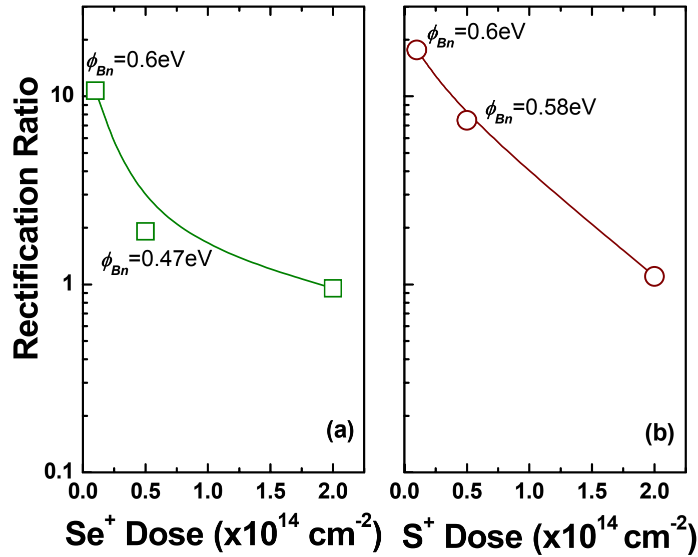


Figure 4.29 The rectification ratio, *RR* improves with increasing Se or S dose. *RR* approaches unity for the samples with the highest dose, 2×10^{14} cm^{-2}. (ϕ_{Bn} is also indicated wherever applicable.)

In this work, *I-V* measurements were carried out at various elevated temperatures (25°C-100°C). This range works well to measure ϕ_B down to 250-200 meV but fails for ϕ_B below 100 meV, and the slope starts to become positive as seen in Figure 4.27. A similar problem has been reported in [13]. Therefore, to accurately characterize the barrier height, temperatures well below room temperature are needed for *I-V* measurements. Instead, rectification ratio, *RR* (defined as I_F/I_R @ |V| = 100mV), is used to evaluate diode rectification characteristics. *RR* is a measure of the ease with which

current flows across a metal-semiconductor junction. For an ideal ohmic contact, *RR*=1. A plot of *RR* as a function of implanted Se and S dose is shown in Figure 4.29. *RR* approaches unity for the samples with the highest dose of Se or S, suggesting corresponding reduction in barrier height to ~0.1 eV levels for both Se and S. This conclusion is consistent with what has been published by other researchers.

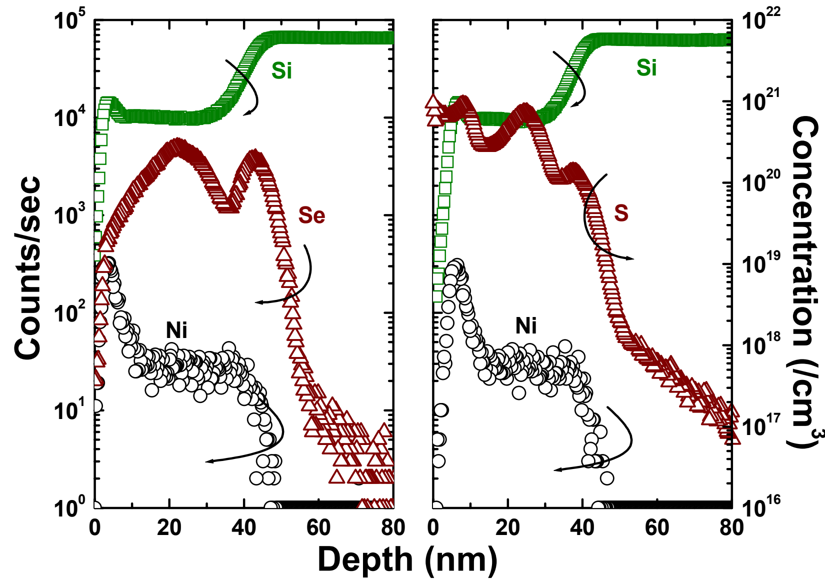


Figure 4.30 The backside SIMS profiles of selenium- and sulfur- implanted NiSi/*n*-Si Schottky diodes. Implanted dose is 2×10^{14} cm^{-2}.

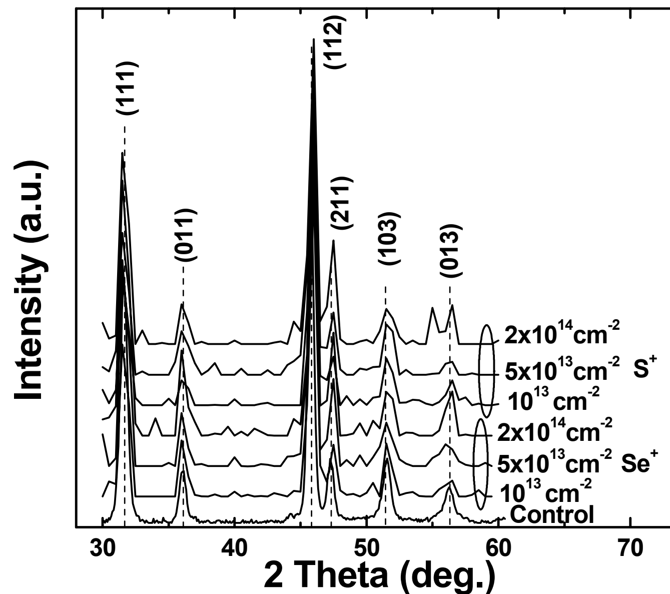


Figure 4.31 GIXRD spectra of selenium and sulfur implanted nickel silicide films.

The backside SIMS analyses further reveal that S and Se are indeed segregated at the NiSi-Si interface as shown in Figure 4.30. The S concentration was calculated using a relative sensitivity factor determined from a standard sample of S implanted in Si. Se was not quantified due to the lack of a standard. The GIXRD data shown in Figure 4.31 also indicates that the incorporation of selenium or sulfur has no effect on the phase of silicide and the desired phase of NiSi is observed for both cases. Furthermore, it is reported that S incorporation does not influence the sheet resistance and thermal stability of NiSi [Zhao 03]. Therefore, a low effective barrier height (~0.1 eV) can be achieved for NiSi on Se (S)-implanted *n*-type Si Schottky diodes without affecting the bulk properties of NiSi.

4.5 Summary

In this chapter, a simple process integration scheme using readily available elements such as nitrogen and fluorine is presented to achieve low barrier height with silicide to *n*-type Si. Wide-range modulation of Schottky barrier height of NiSi/*n*-type Si contact is successfully demonstrated with nitrogen implantation. However, the change in electron barrier height is not matched by corresponding change in hole barrier height at higher implanted doses. This mismatch is explained by the effective reduction in the silicon bandgap due to strain. The effects of nitrogen on the resistivity of nickel silicide become more pronounced beyond a certain implanted dose. Fluorine has little effect on the barrier height of NiSi/*n*-type Si contact.

The effective electron barrier heights achieved with the various species investigated in this work are summarized in Figure 4.32. Sub-0.1 eV barrier height is obtained with the use of selenium and sulfur, but their impact on CMOS process flow is

not well understood. Nitrogen, in contrast, presents a highly manufacturable solution to achieve low barrier contact technology.

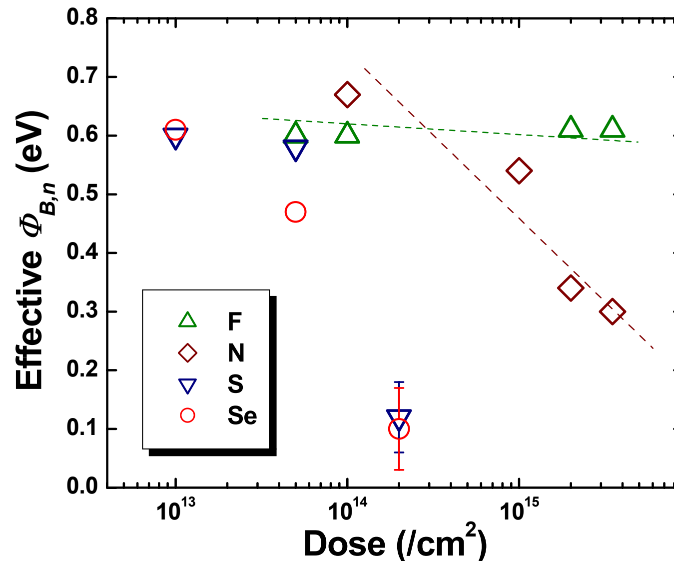


Figure 4.32 Summary of effective electron barrier heights of NiSi/n-type lightly doped Si diodes with the various species studied in this work. These species are implanted into the silicon prior to silicidation and are then segregated to the silicide-Si interface.

4.6 References

[1] International Technology Roadmap for Semiconductors, 2007 Edition, (http://www.public.itrs.net/).

[2] S. -D. Kim, C. -M. Park, and J. C. S. Woo, "Advanced Model and Analysis of Series Resistance for CMOS Scaling Into Nanometer Regime—Part II: Quantitative Analysis", *IEEE Transactions on Electron Devices*, vol. 49, no. 3, pp. 467-472, 2002.

[3] K. Mistry, C. Allen, C. Auth, B. Beattie, D. Bergstrom, M. Bost, M. Brazier, M. Buehler, A. Cappellani, R. Chau, C. -H. Choi, G. Ding, K. Fischer, T. Ghani, R. Grover, W. Han, D. Hanken, M. Hattendorf, J. He, J. Hicks, R. Huessner, D. Ingerly, P. Jain, R. James, L. Jong, S. Joshi, C. Kenyon, K. Kuhn, K. Lee, H. Liu, J.

Maiz, B. Mclntyre, P. Moon, J. Neirynck, S. Pae, C. Parker, D. Parsons, C. Prasad, L. Pipes, M. Prince, P. Ranade, T. Reynolds, J. Sandford, L. Shifren, J. Sebastian, J. Seiple, D. Simon, S. Sivakumar, P. Smith, C. Thomas, T. Troeger, P. Vandervoorn, S. Williams, and K. Zawadzki, " A 45nm Logic Technology with High-k+Metal Gate Transistors, Strained Silicon, 9 Cu Interconnect Layers, 193nm Dry Patterning, and 100% Pb-free Packaging," *IEEE IEDM Tech. Digest*, pp. 247–250, 2007.

[4] P. Packan, "Device and Circuit Interactions," *IEEE IEDM Short Course*, 2007.

[5] A. Yagishita, T. -J. King, and J. Bokor, "Schottky Barrier Height Reduction and Drive Current Improvement in Metal Source/Drain MOSFET with Strained-Si Channel," *J. J. Applied Physics*, vol. 43, no. 4B, pp. 1713-1716, 2004.

[6] J. Liu, and M. C. Ozturk, "Nickel Germanosilicide Contacts Formed on Heavily Boron Doped $Si_{1-x}Ge_x$ Source/Drain Junctions for Nanoscale CMOS," *IEEE Transactions on Electron Devices*, vol. 52, no. 7, pp. 1535-1540, 2005.

[7] J. Kedzierski, P. Xuan, E. H. Andew, J. Bokor, T.-J. King, and C. Hu, "Complementary Silicide Source/Drain Thin-body MOSFETs for the 20nm Gate Length Regime," *IEEE IEDM Tech. Digest*, pp. 57-60, 2000.

[8] M. Tao, S. Agarwal, D. Udeshi, N. Basit, E. Maldonado, and W. P. Kirk, "Low Schottky Barriers on n-type Silicon (001)," *Applied Physics Letters*, vol. 83, no. 13, pp. 2593-2595, 2003.

[9] D. Connelly, C. Faulkner, D. E. Grupp, and J. S. Harris, "A New Route to Zero-Barrier Metal Source/Drain MOSFETs," *IEEE Transactions on Nanotechnology*, vol. 3, no. 1, pp. 98-104, 2004.

[10] J. M. Shannon, "Reducing the Effective Height of a Schottky Barrier Using Low-energy Ion Implantation," *Applied Physics Letters*, vol. 24, no. 8, pp.369-371, 1974.

[11] A. Kinoshita, Y. Tsuchiya, A. Yagishita, K. Uchida and J. Koga, "Solution for High-Performance Schottky-Source/Drain MOSFETs: Schottky Barrier Height Engineering with Dopant Segregation Technique," *IEEE VLSI Technology Symp. Tech. Digest*, pp. 168-169, 2004.

[12] M. Zhang, J. Knoch, Q.T. Zhao, St. Lenk, U. Breuer and S. Mantl, " Schottky Barrier Height Modulation Using Dopant Segregation in Schottky-barrier SOI-MOSFETs," *ESSDERC Proc.*, pp. 457-460, 2005.

[13] Q. T. Zhao, U. Breuer, E. Rije, St. Lenk, and S. Mantl, "Tuning of NiSi/Si Schottky Barrier Heights by Sulfur Segregation during Ni Silicidation," *Applied Physics Letters*, vol. 86, no. 6, pp. 062108, 2005.

[14] H. -S. Wong, L. Chan, G. Samudra, and Y. -C. Yeo, "Effective Schottky Barrier Height Reduction Using Sulfur or Selenium at the NiSi/n-Si (100) Interface for Low Resistance Contacts," *IEEE Electron Device Letters*, vol. 28, no. 12, pp. 1102-1104, 2007.

[15] K. Ikeda, Y. Yamashita, N. Sugiyama, N. Taoka, and S. Takagi, "Modulation of NiGe/Ge Schottky Barrier Height by Sulfur Segregation During Ni Germanidation," *Applied Physics Letters*, vol. 88, no. 15, pp. 152115, 2006.

[16] R. T. P. Lee, T.-Y. Liow, K.-M. Tan, A. E.-J. Lim, H.-S. Wong, P.-C. Lim, D. M. Y. Lai, G.-Q. Lo, C.-H. Tung, G. Samudra, D.-Z. Chi, and Y.-C. Yeo, "Novel Nickel-Alloy Silicides for Source/Drain Contact Resistance Reduction in N-

Channel Multiple-Gate Transistors with Sub-35nm Gate Length," *IEEE IEDM Tech. Digest*, pp. 851-854, 2006.

[17] Y. Nishi, Y. Tsuchiya, A. Kinoshita, T. Yamauchi, and J. Koga, "Interfacial Segregation of Metal at NiSi/Si Junction for Novel Dual Silicide Technology," *IEEE IEDM Tech. Digest*, pp. 135–138, 2007.

[18] H. N. Chern, C. L. Lee, and T. F. Lei, "The Effects of Fluorine Passivation on Polysilicon Thin-Film Transistors," *IEEE Transactions on Electron Device*, vol. 41, no. 5, pp. 698-702, 1994.

[19] C. K. Yang, T. F. Lei, and C. L. Lee, "Characteristics of Top-Gate Thin-Film Transistors Fabricated on Nitrogen-Implanted Polysilicon Films," *IEEE Transactions on Electron Devices,* vol. 42, no. 12, pp. 2163–2169, 1995.

[20] D. A. Good, "Novel Processes for Poly-Si Thin-Film Transistors on Plastic Substrates," Ph.D. dissertation, University of California, Berkeley, Department of Elec. Engg. and Computer Sciences, 2007.

[21] E. G. Colgan, J. P. Gambino, and Q. Z. Hong, "Formation and Stability of Silicides on Polycrystalline Silicon," *Mater. Sci. Eng*. R16, pp. 43-96, 1996.

[22] Power Diffraction Standards, ICDD, Card 38-0833 (2002).

[23] J. P. Gambino and E. G. Colgan, "Silicides and Ohmic Contacts," *Materials Chemistry and Physics*, vol. 52, pp. 99-146, 1998.

[24] S. -L. Zhang and M. Östling, "Metal Silicides in CMOS Technology: Past, Present, and Future Trends," *Critical Reviews Solid State Material Sciences*, vol. 28, no. 1, pp. 1-129, 2003.

[25] Q. T. Zhao, E. Rije, U. Breuer, St. Lenk, and S. Mantl, "Tuning of Silicide Schottky Barrier Heights by Segregation of Sulfur Atoms," *ICSSICT Proc.*, pp. 456-459, 2004.

Chapter 5

Contact Material for Ge MOSFETs

5.1 Introduction

For high-performance logic, the performance metric is the MOSFET intrinsic delay, $\tau = CV/I$, where C is the total gate capacitance (including parasitic gate overlap and fringing capacitance) per transistor width, V is the power supply voltage, and I is the saturation drive current per transistor width. To keep pace with the historical rate of improvement in transistor performance, the target is to achieve a 17% per year increase in $1/\tau$ as shown in Figure 5.1 [1]. In a bulk Si MOSFET, gate dielectric scaling has been the key enabler to reduce τ because a thinner gate dielectric improves drive current. The MOSFET delay can also be decreased by raising drive current via increasing the channel mobility as discussed next.

The saturation drive current ($I_{d,sat}$) of a very short-channel MOSFET is believed to be limited by carrier injection from the source into the channel and is written as [2] –

$$I_{dsat} = W Q_i(x_0) v_{xo} \tag{5.1}$$

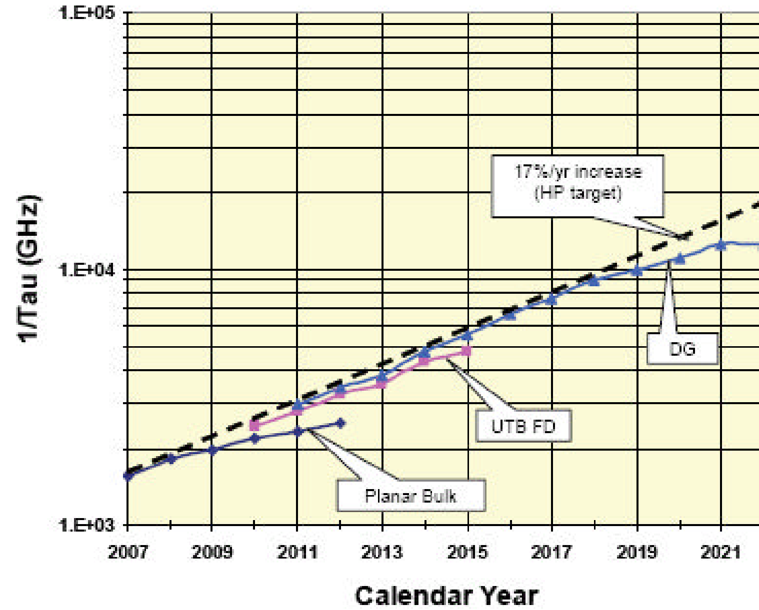


Figure 5.1 Projected scaling of transistor intrinsic speed (1/τ) for high-performance (HP) logic [1]. For planar bulk-Si technology, the 1/τ curve is projected to fall short of the desired 17%/year increase target by 2009. Implementation of ultra-thin body fully depleted (UTB FD) SOI MOSFETs and double-gate (DG) MOSFETs will help to exceed 2009 projections for 1/τ but only until 2019.

where W is the channel width, $Q_i(x_0)$ is the channel areal density at the virtual source and v_{xo} is the carrier velocity at the virtual source. Virtual source position (x_0) in the channel is shown in Figure 5.2. Carrier velocity at the virtual source can be written as –

$$v_{xo} = \left(\frac{1-r}{1+r}\right) v_\theta \tag{5.2}$$

where r is the backscattering coefficient and v_θ is the unidirectional thermal velocity. Both these factors strongly depend on the low-field mobility near the source edge. As a consequence, higher low-field mobility is needed to enhance the saturation drive current in future MOSFETs.

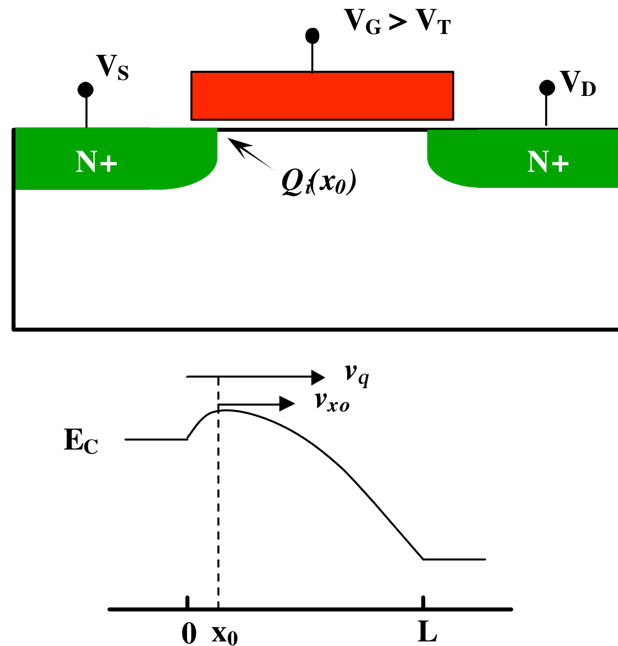


Figure 5.2 Virtual source location (x_0) in the channel.

The first potential solution is to enhance mobility by strain engineering. Several approaches have been reported in the literature to implement enhanced mobility via strain incorporation – (a) at the wafer level [3, 4], (b) at the package level, or (c) by novel processes [5-7].

In theory, the highest performance can result from a combination of strain engineering and the introduction of novel high mobility channel materials. Germanium (Ge) has been identified as a potential candidate to replace silicon for *p*-channel devices due to its higher hole mobility (4× relative to Si). The lack of a stable dielectric on the germanium channel has long been seen as the major limitation for the fabrication of Ge MOSFETs. However, with recent advancements, Ge MOSFETs with high-κ gate dielectric are gaining a lot of attention for highly-scaled high performance technology [8-

11]. Higher bulk mobility and a demonstrated capability with high-κ dielectrics make Ge a potential candidate for improving PMOSFET drive current while allowing for continued aggressive scaling of the gate oxide.

5.2 Heteroepitaxial Growth of Ge on Silicon

Due to the large lattice mismatch (4%) between Ge and Si, it is difficult to grow high-quality Ge layers on silicon. This large mismatch is responsible for biaxial compression in epitaxially grown Ge layers. As the epitaxial film grows thicker, strain is relaxed by the generation of misfit dislocations. The thickness below which the epitaxial layer remains fully strained is defined as the critical thickness, t_c. Thus for strained layers, t_c is an important device design parameter. t_c depends strongly on the growth temperature and growth mechanism. Figure 5.3 shows the critical thickness for $Si_{1-x}Ge_x$ grown on Si, as a function of Ge mole fraction [12].

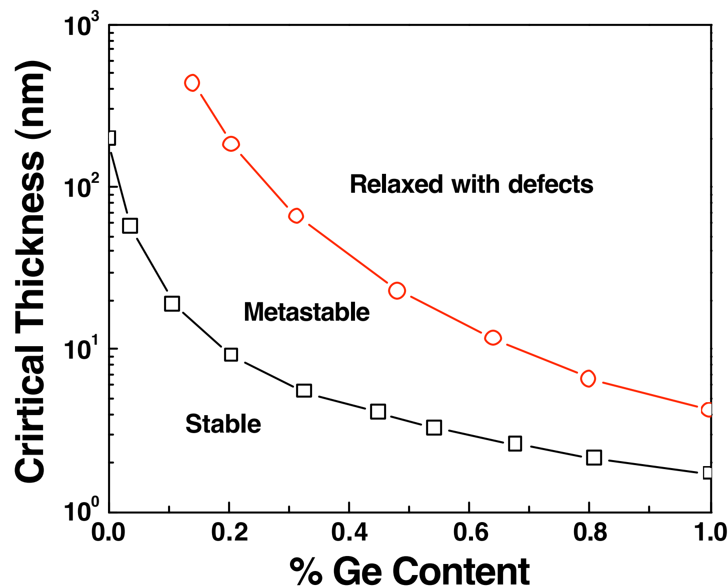


Figure 5.3. Critical thickness as a function of the Ge mole fraction (adapted from [12]).

The stable zone indicates the regime for which no dislocations will be formed during growth. As the layer thickness is increased, growth enters the metastable regime for which no dislocations are formed during growth, but relaxation may occur upon subsequent heat treatments. For even thicker films, dislocations are formed during growth and strain is relaxed. Note that t_c decreases rapidly as the Ge mole fraction increases, as shown in Figure 5.3. For pure Ge, the critical thickness is about 10-15 Å.

There are three modes of heteroepitaxial growth : Frank-van der Merwe (F-vdM), Volmer-Weber (V-W), and Stranski-Krastanow (S-K). F-vdM is layer-by-layer growth (2D), V-W is island growth (3D), and S-K proceeds initially as layer-by-layer, followed by islands formation.

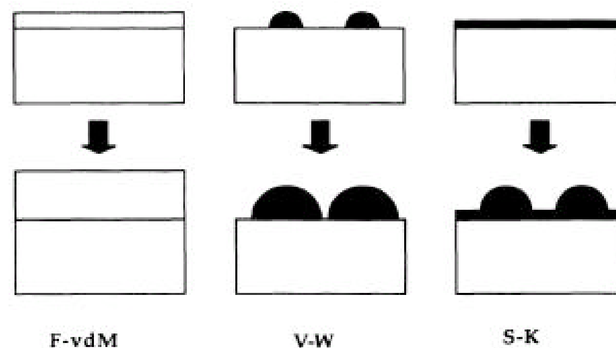


Figure 5.4. Schematic diagram of three possible growth modes: Frank-van der Merwe (F-vdM), Volmer-Weber (V-W), and Stranski-Krastanow (S-K) (adapted from [13]).

The growth of Ge on Si is S-K [13] and layer-to-island transition occurs at sub-1 nm thickness for Ge-rich films. However, by employing low substrate temperature (300-375°C), it is possible to suppress the formation of islands and the layer-by-layer growth mode can be achieved over extended thickness ranges because the growth rate is now surface-reaction limited [14].

5.3 Metal Germanides

To fully exploit the superior carrier transport properties of Ge, focused research on developing self-aligned germanides to contact Ge MOSFETs needs to be explored, in analogy with current Si-based technology where a self-aligned silicide (salicide) process is used to contact the source, drain, and gate. Additionally, germanide should exhibit low sheet resistance, good thermal stability of the lowest resistance phase, and lower formation temperature. In order to obtain a self-aligned feature, germanide should be resistant to the chemical etching used to remove unreacted metal. Several different germanides — titanium germanide, cobalt germanide and nickel germanide — have been reported in literature. Titanium germanide forms at high temperature (~800°C) and exhibits a resistivity of 20 μΩ–cm [15]. The formation of cobalt germanide occurs at a relatively lower temperature (~425°C) but shows high resistivity (35 μΩ–cm) [15]. Nickel germanide appears to be the most promising candidate because it is formed at a lower temperature (~ 150°C), and is characterized by the sheet resistivity of ~15-20 μΩ–cm [16-20].

All of the aforementioned studies were done on bulk germanium wafers. Due to their poor mechanical strength and high cost, bulk germanium wafers are not compatible with CMOS processing. Much progress has been made toward integrating high-quality, low-defect density epi-Ge layers on silicon substrates, thus making germanium channel devices attractive for high-volume manufacturing. The goal of this study is to investigate material properties of nickel germanide formed on epi-Ge on silicon substrates.

5.4 Experimental Details

For the epi-Ge layer on silicon substrate, a 120 nm thick high-quality Ge epitaxial film was grown using rapid thermal chemical vapor deposition. The substrate temperature was in the range of 330–370°C. As described in section 5.2, low temperature processing suppresses island formation, thus improving surface roughness and the quality of grown films. The epi-Ge layer is fully relaxed because it is thicker than the critical thickness.

Figure 5.5 shows a high-resolution transmission electron microscope (HRTEM) image of epi-Ge film grown on Si [21]. For this particular image, ZrO_2 and TaN were also deposited after epi-Ge growth. The threading dislocation density of the Ge epitaxial layer is too low to be seen by HRTEM. The secondary ion mass spectroscopy (SIMS) analysis result of epi-Ge film on silicon is shown in Figure 5.6.

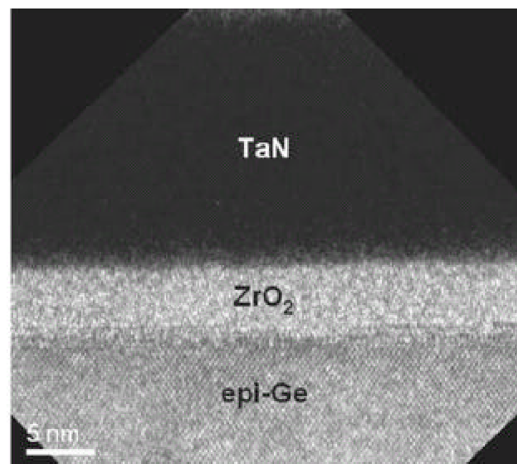


Figure 5.5 HR XTEM image of relaxed epi-Ge film on silicon substrate [21].

A RF-sputtering system was used to deposit 40 nm-thick nickel after epi-Ge growth. A TiN capping layer was also deposited on some samples. Next, a nickel germanidation process was carried out using rapid thermal anneal (RTA) in nitrogen

ambient for 60 seconds. The temperature of germanidation anneal was varied from 300°C to 400°C. The complete process flow is illustrated in Figure 5.7.

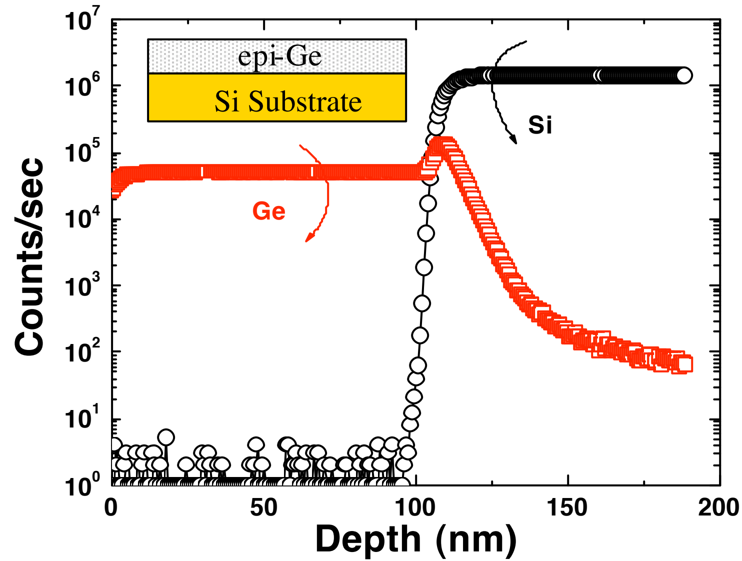


Figure 5.6 SIMS profiles obtained for the epi-Ge films on silicon substrate.

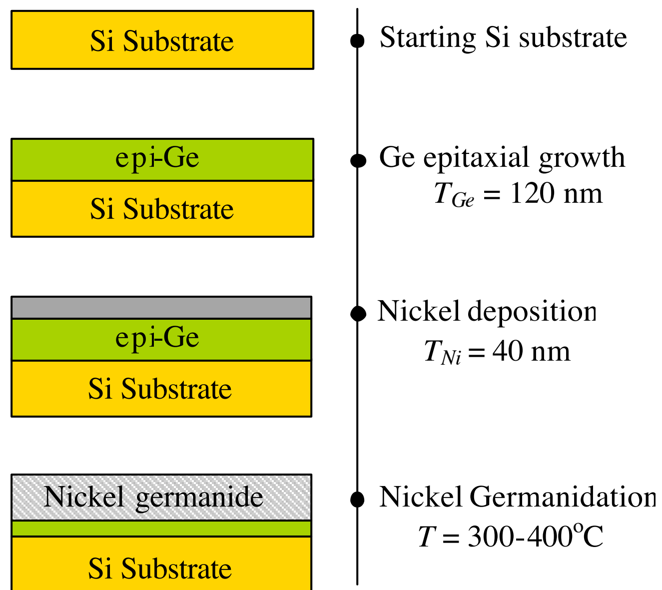


Figure 5.7 Process flow used in this study to investigate nickel germanide film properties formed on epi-Ge on silicon substrate.

Phase identification of thus formed nickel germanide films was carried out by grazing incidence X-ray diffraction (GIXRD). Scanning electron microscopy (SEM) was done for visual examination. A four-point probe was used to measure sheet resistance of germanide films.

5.5 Physical Analysis

GIXRD has been performed ex-situ using a diffractometer equipped with a temperature chamber and a Cu tube. Figure 5.8 shows XRD spectra for annealed germanide samples. For all the samples, peaks at 34.6°, 42.7°, 44.2°, 45.6°, and 53.5° are observed. Peak positions are in good agreement with nickel mono-germanide (NiGe) phase and are identified as NiGe (111), (210), (112), (211), and (212) peaks respectively. No other peaks corresponding to Ni_2Ge or Ni_5Ge_3 are observed. It is widely believed that the Ni-Ge phase sequence is the formation of Ni_2Ge [22] or Ni_5Ge_3 [23] at lower temperature followed by the formation of NiGe at ~200°C. These results suggest that the complete formation of NiGe has occurred in the annealing temperature range of the investigation.

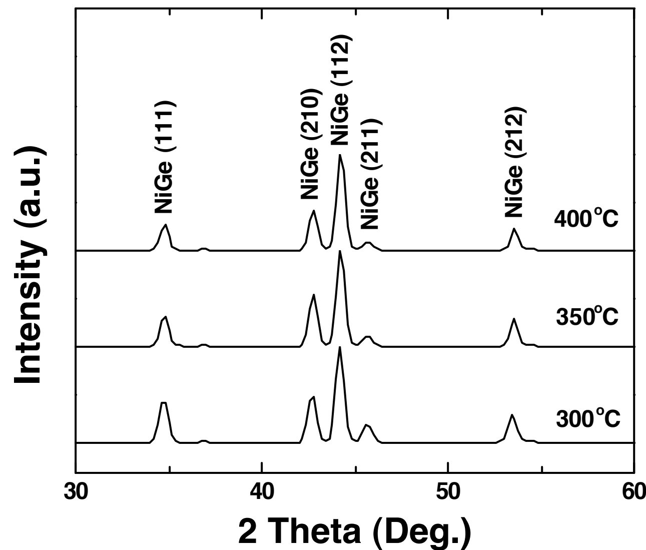


Figure 5.8 GIXRD Spectra of nickel germanide films.

Cross-sectional SEM results indicate the thickness of formed germanide at various temperatures as shown in Figure 5.9. Polycrystalline NiGe with large grain size (~0.1 μm) is observed in all the samples. The interface between individual grain and germanium is smooth, but germanide thickness variation is ~200 Å. Severe grain boundary grooving is also seen. Average germanide thickness of 100 nm is seen in every annealed sample. The ratio of consumed Ni to formed NiGe thickness is ~ 1:2.5, which is consistent with other published work.

Figure 5.10 shows measured sheet resistance of germanide films using a four-point probe. After annealing to 300°C, the sheet resistance of 1.65 Ω is recorded. This is

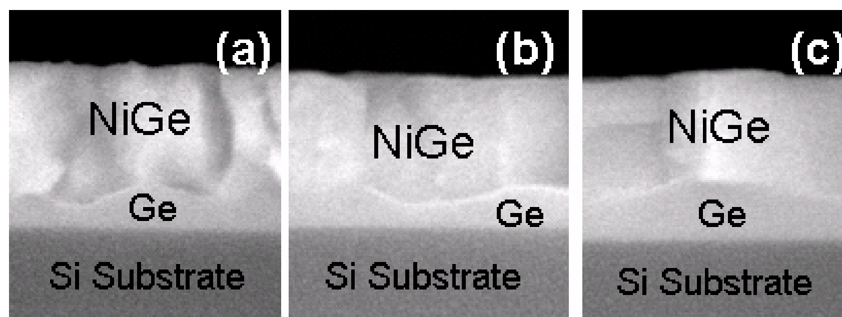


Figure 5.9 Cross-section SEM of formed germanide at various temperatures – (a) 300°C, (b) 350°C, (c) 400°C.

equivalent of 16 μΩ-cm sheet resistivity, confirming the formation of the lowest resistivity nickel germanide phase. With further increase in annealing temperature to 350°C, sheet resistivity is observed to remain constant, but increasing temperature to 400°C results in resistivity of 30 μΩ-cm, an increase of 1.8×. Atomic force microscopy (AFM) confirms the presence of a larger grain size for the 400°C annealed sample. Thus, the increase in resistivity is attributed to a morphological change in NiGe at 400°C due to increased grain grooving and/or increased roughness.

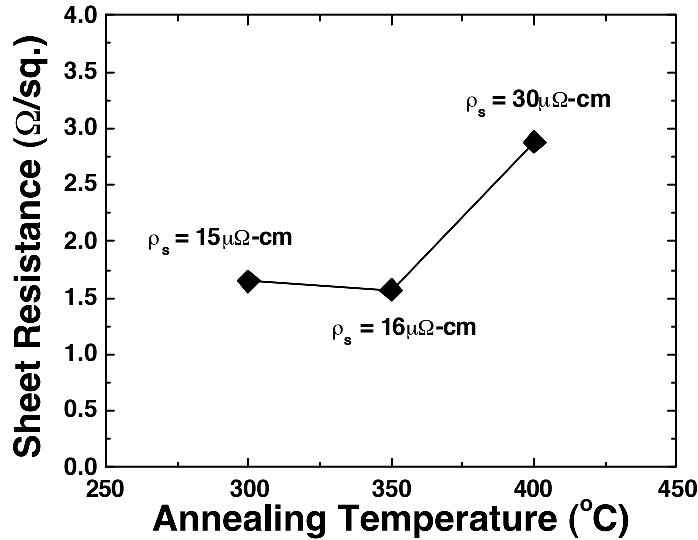


Figure 5.10 Sheet resistance of nickel germanide *vs*. annealing temperature.

5.6 Wet Chemical Reaction

The self-aligned silicide (salicide) process requires an etchant that has a high etch rate for the unreacted metal and a low etch rate for the silicide. In addition, the etchant used for metal removal shouldn't attack the gate-sidewall spacer material. For selective removal of unreacted nickel over nickel germanide, first, the use of classical wet chemicals conventionally known for removing unreacted nickel selectively in corresponding nickel salicide process was attempted.

Most of the well-known chemistries used for nickel removal—SPM (H_2SO_4: H_2O_2), $FeCl_3$—etch nickel germanide as well. This observation can be correlated to the chemical reaction mechanism of nickel germanide with SPM (and $FeCl_3$). Both agents etch nickel and nickel germanide without any selectivity because the main reaction is happening between Ni and the chemical agent. Ni-Ge bonding is not strong enough to become inert to the chemical agents. Since only one germanide phase (NiGe) was obtained in all the samples, no etch rate variation was seen. It is also observed that SC1

(DI water: H_2O_2: NH_4OH) could etch nickel germanide if germanide was capped with titanium nitride (TiN). However, if no TiN capping layer was used, SC1 did not etch nickel germanide or nickel. In SC1, H_2O_2 promotes the formation of metal oxide and NH_4OH slowly etches those oxides. The aforementioned experimental observation can thus be explained by Ti diffusion into nickel germanide, which helps expedite the wet etch process because Ti-alloys are generally very susceptible to an oxidizing ammoniacal SC1 wet etch process. Irrespective of use of the capping layer, selective removal of unreacted nickel is not achieved with SC1.

Sulfuric acid (H_2SO_4:H_2O) reacts with most metals including nickel in a single displacement reaction to produce H_2 gas and the metal sulfate. With careful optimization of the chemical condition (20% H_2SO_4), it is possible to selectively remove nickel over nickel germanide because 20% H_2SO_4 is not strong enough to break Ni-Ge bonds to form nickel sulfate. This process was verified with different time conditions and no loss of formed nickel germanide was observed in all studied samples. On the other hand, unreacted nickel is etched in a very short time. This demonstration shows that the developed wet etch process can remove unreacted nickel with high selectivity over nickel germanide. Cross-sectional SEM images of nickel and nickel germanide before and after etching in different chemicals are shown in Figures 5.11 and 5.12.

Furthermore, commonly known spacer material silicon nitride (Si_3N_4) and tetra ethyl ortho-silicate (TEOS) oxide are immune to the developed process for selective removal of nickel. Similar wet chemical treatment was given to all the samples formed at different temperatures and it is observed that selectivity is maintained over the investigated temperature range.

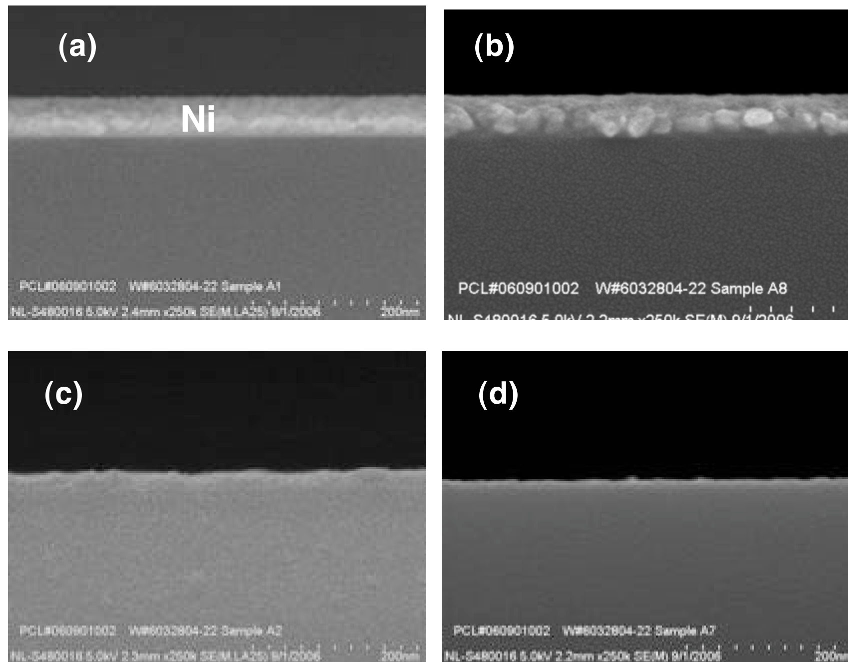


Figure 5.11 Etching characteristics of nickel film in SC1, SPM and 20% H_2SO_4: (a) sputtered nickel film on Si substrate; (b) after exposure to SC1; (c) after exposure to SPM; (d) after exposure to 20% H_2SO_4.

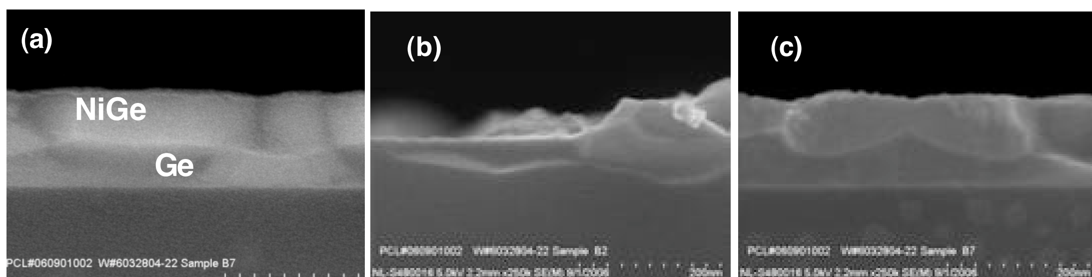


Figure 5.12 Cross-sectional SEM images after NiGe etching in SPM and 20% H_2SO_4: (a) NiGe film after germanidation, (b) after exposure to SPM, (c) after exposure to 20% H_2SO_4. Thus 20% H_2SO_4 can be successfully used to remove nickel with high selectivity.

5.7 Summary

As the possibility of using a Ge channel arises, it becomes important to investigate the contact material for epi-Ge on Si devices. This study demonstrates that nickel germanide is a potentially attractive candidate contact material for achieving low

resistance and low thermal budget. To realize the self-aligned feature, a wet chemical treatment using H_2SO_4 is developed. It is demonstrated that unreacted nickel can be successfully removed with very high selectivity over nickel germanide or spacer material.

5.8 References

[1] International Technology Roadmap for Semiconductors, 2007 Edition, (http://www.public.itrs.net/).

[2] M. Lundstrom, "Elementary Scattering Theory of the Si MOSFET," *IEEE Electron Device Letters*, vol. 21, no. 7, pp. 361-363, 1997.

[3] J. Welser, J. L. Hoyt, and J. F. Gibbons, "NMOS and PMOS Transistors Fabricated in Strained Silicon/Relaxed Silicon-germanium Structures," *IEEE IEDM Tech. Digest*, pp. 1000-1002, 1992.

[4] K. Rim, J. Chu, K. A. Jenkins, T. Kanarshy, K. Lee, A. Mocuta, H. Zhu, R. Roy, J. Newbury, J. Ott, K. Petrarca, P. Mooney, D. Lacey, S. Koester, K. Chan, D. Boyd, M. Ieong, and H.-S. P. Wong, "Characteristics and Device Design of Sub-100 nm Strained-Si N- and P-MOSFETs," *IEEE VLSI Technology Symp. Tech. Digest*, pp. 98-99, 2002.

[5] S. Ito, H. Namba, K. Yamaguchi, T. Hirata, K. Ando, S. Koyama, S. Kuroki, N. Ikezawa, T. Suzuki, T. Saitoch, and T. Horiuchi, "Mechanical Stress Effect of Etch-stop Nitride and its Impact on Deep Submicron Transistor Design," *IEEE IEDM Tech. Digest*, pp. 247-250, 2000.

[6] A. Shimizu, K. Hachimin, N. Ohki, H. Ohta, M. Koguchi, Y. Nonaka, H. Sato, and F. Ootsuka, "Local Mechanical-Stress Control (LMC): A New Technique for CMOS Performance Enhancement," *IEEE IEDM Tech. Digest*, pp. 433-437, 2001.

[7] T. Ghani, M. Armstrong, C. Auth, M. Bost, P. Charvat, G. Glass, T. Hoffmann, K. Johnson, C. Kenyon, J. Klaus, B. McIntyre, K. Mistry, A. Murthy, J. Sandford, M. Silberstein, S. Sivakumar, P. Smith, K. Zawadzki, S. Thompson, and M. Bohr, "A 90 nm High Volume Manufacturing Logic Technology Featuring Novel 45 nm Gate Length Strained Silicon CMOS Transistors," *IEEE IEDM Tech. Digest*, pp. 978–980, 2003.

[8] C. O. Chui, H. Kim, D. Chi, B. B. Triplett, P. C. McIntyre, and K. C. Saraswat, "A Sub-400°C Germanium MOSFET Technology with High-κ Dielectric and Metal Gate," *IEEE IEDM Tech. Digest*, pp. 437-440, 2002.

[9] C. H. Huang, M. Y. Yang, A. Chin, W. J. Chen, C. X. Zhu, B. J. Cho, M.-F. Li, and D. L. Kwong, "Very Low Defects and High Performance Ge-On-Insulator p-MOSFETs with Al_2O_3 Gate Dielectrics," *IEEE VLSI Technology Symp. Tech. Digest*, pp. 119-120, 2003.

[10] W. P. Bai, N. Lu, J. Liu, A. Ramirez, D. L. Kwong, D. Wristers, A. Ritenour, L. Lee, and D. Antoniadis, "Ge MOS Characteristics with CVD HfO_2 Gate Dielectrics and TaN Gate Electrode," *IEEE VLSI Technology Symp. Tech. Digest*, pp. 121-122, 2003.

[11] P. Zimmerman, G. Nicholas, B. De Jaeger, B. Kaczer, A. Stesmans, L.-A. Ragnarsson, D. P. Brunco, F. E. Leys, M. Caymax, G. Winderickx, K. Opsomer, M. Meuris, and M. M. Heyns, "High Performance Ge pMOS Devices Using a Si-Compatible Process Flow," *IEEE IEDM Tech. Digest*, pp. 437-440, 2006.

[12] J. W. Matthews and A. E. Blakeslee, "Defects in Epitaxial Multilayers I. Misfit Dislocations," *J. Crystal Growth*, vol. 27, pp. 118-125, 1974.

[13] D. J. Eaglesham and M. Cerullo, "Dislocation-Free Stranski-Krastanow Growth of Ge on Si (100)," *Physical Review Letters*, vol. 64, no. 16, pp. 1943-1946, 1990.

[14] B. Cunningham, J. O. Chu, and S. Akbar, "Heteroepitaxial Growth of Ge on (100) Si by Ultrahigh Vacuum, Chemical Vapor Deposition," *Applied Physics Letters*, vol. 59, pp. 3574-3576, 1991.

[15] S. P. Ashburn, M. C. Ozturk, J. J. Wortman, G. Harris, J. Honeycutt, and D. M. Maher, "Formation of Titanium and Cobalt Germanides on Si(100) Using Rapid Thermal Processing," *J. Electronic Materials*, vol. 21, no. 1, pp. 81-86, 1992.

[16] J. Y. Spann, R. A. Anderson, T. J. Thornton, G. Harris, S. G. Thomas, and C. Tracy, "Characterization of Nickel Germanide Thin Films for Use as Contacts to p-Channel Germanium MOSFETs," *IEEE Electron Device Letters*, vol. 26, no. 3, pp.151-153, 2005.

[17] S. Zhu, A. Nakajima, Y. Yokoyama, and K. Ohkura, "Temperature Dependence of Ni-Germanide Formed by Ni-Ge Solid-State Reaction" *Ext. Abs. Int. Workshop Junction Technology*, pp. 89-92, 2005.

[18] Q. Zhang, N. Wu, T. Osipowicz, L. K. Bera, and C. Zhu, "Formation and Thermal Stability of Nickel Germanide on Germanium Substrate," *Jpn. J. Applied Physics*, vol. 44, no. 45, pp. L1389-L1391, 2005.

[19] H. B. Yao, C. C. Tan, S. L. Liew, C. T. Chua, C. K, Chua, R. Li, R. T. P. Lee, S. J. Lee, and D. Z. Chi, "Material and Electrical Characterization of Ni- and Pt-Germanides for p-channel Germanium Schottky Source/Drain Transistors," *Int. Workshop Junction Technology,* pp.164-169, 2006.

[20] F. Nemouchi, D. Mangelinck, C. Bergman, G. Clugnet, P. Gas, and J. L. Lábár, "Simultaneous Growth of Ni_5Ge_3 and NiGe by Reaction of Ni Film with Ge," *Applied Physics Letters*, vol. 89, pp. 131920, 2006.

[21] J. Oh, P. Majhi, C. Y. Kang, J.-W. Yang, H. -H. Tseng, and R. Jammy, "Thermal Stability of Nanoscale Ge Metal-Oxide-Semiconductor Capacitors with ZrO_2 High-*k* Gate Dielectrics on Ge Epitaxial Layers," *Applied Physics Letters*, vol. 90, pp. 202102, 2007.

[22] Y. F. Hsieh, L. J. Chen, E. D. Marshall, and S. S. Lau, "Partial Epitaxial Growth of Ni_2Ge and NiGe on Ge(111), " *Thin Solid Films*, vol. 162, pp. 287-294, 1988.

[23] J. K. Patterson, B. J. Park, K. Ritley, H. Z. Xiao, L. H. Allen, and A. Rockett, "Kinetics of Ni/a-Ge Bilayer Reactions," *Thin Solid Films*, vol. 253, no. 1-2, pp. 456-461, 1994.

Chapter 6

Conclusions

6.1 Summary

After more than 40 years of aggressive downsizing of MOSFETs, geometric scaling is reaching fundamental limits. New materials and novel processes will be necessary in order to extend device scaling to the last CMOS technology node of the *International Technology Roadmap for Semiconductors* (ITRS). Process-induced strain has already become an industry standard since 90 nm logic technology node to boost mobility in the channels, and traditional gate stacks based on SiO_xN_y/poly-Si are now being replaced by high-κ/metal gate for the leading 45 nm logic technology (as gate leakage becomes too high to manage). For sub-45 nm CMOS generations, the most pressing challenge is to reduce the parasitic series resistance associated with source/drain structures. There are several components of series resistance, but the primary and the dominant ones of concern are related to ultra-shallow junctions (S/D extension resistance) and silicide contacts to the deep source/drain regions (contact resistance).

Historically, thinning down the gate dielectric and increasing channel doping have been widely implemented to suppress short channel effects to support the gate length shrink. However, the gate dielectric scaling has slowed down, and even with the adoption of high-κ/metal-gate it will be difficult to scale *EOT* well below 1 nm. The increase in channel doping leads to degraded mobility, band-to-band tunneling and device-to-device variations due to random dopant fluctuations. These two physical limitations necessitate the formation of ultra-shallow junctions (USJs) to maintain good electrostatic integrity for sub-45 nm nodes. The downside to this is the increase in the S/D extension resistance. Due to continued scaling of the contacted pitch (and thus S/D contact area), the contact resistance is becoming a major issue and will severely degrade MOSFET performance as CMOS is scaled into the nano-scale regime. This problem is exacerbated for advanced device structures which use thin silicon on insulator.

6.2 Contributions

This dissertation has provided an outlook of the aforementioned issues associated with source/drain design, an overview of the state-of-the-art,and opportunities that lie ahead to keep pace with the historical performance improvement rate. These issues are addressed with the aim of limiting the impact of series resistance on device performance. Along these lines, successful experimental demonstrations of possible solutions are presented: (a) formation of low resistance USJs with high active dopant concentrations for controlled short channel effects (SCE), (b) integration of USJ process with high-κ/metal gate on Si platform, (c) low Schottky barrier of a silicide to *n*-type Si junction for reduced contact resistance. The major conclusions of the topics listed above are summarized as follows.

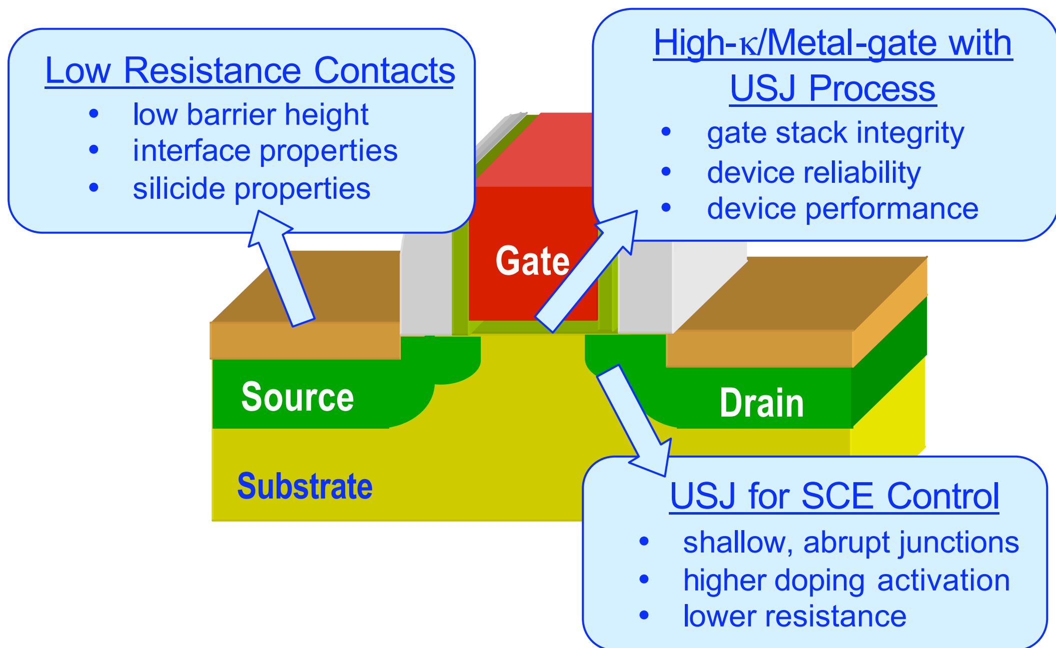


Figure 6.1. Source/Drain engineering issues addressed in this dissertation.

Through a systematic study, the formation of USJs is investigated by conventional and advanced processing methods. These results indicate that in order to meet the requirements for sub-45 nm CMOS generations, spike annealing will not be adequate, especially for *p+/n* junctions due to boron transient enhanced diffusion. Flash annealing, in contrast, appears to be a promising alternative because it provides for shorter annealing time (to limit dopant diffusion) and higher peak temperature (to maximize dopant activation). Sub-15 nm junctions with lower sheet resistance (~1000Ω/€) were achieved with flash annealing, meeting the specifications for 32 nm CMOS technology node. Furthermore, fluorine co-implantation has been shown to reduce the junction leakage of the flash-annealed samples by a significant factor. The advantages of GCIB doping

(infusion doping) over conventional ion implantation are also presented and it shows no evidence of channeling for either (100)- or (110)-oriented Si substrates. Therefore, GCIB doping is attractive for fabrication of ultra-shallow junctions for (110) pMOSFETs, if boron channeling is the limiting factor. 32 nm node USJ requirements are also met by infusion doped, flash annealed junctions.

In translating USJs results to devices in the technologies where flash annealing technique would be used, the integration of flash annealing in the CMOS fabrication process flow becomes an important question. With recent reports showing that transistors with high-κ and metal-gates have been introduced into 45 nm node, the effects of flash annealing on the performance and reliability of MOSFETs with high-κ (HfSiON)/metal-gate (TiN) stack are presented primarily to investigate potential roadblocks (if any) associated with the process integration of the two. A conventional gate-first CMOS process flow was used to fabricate devices with sub-100 nm gate length. It is shown that the flash annealing process has minimal effects on gate-stack properties such as *EOT*, T_{phys}, V_{FB}, and leakage, so it appears to be compatible with high-κ/metal-gate stacks while achieving the benefits of USJs (suppression of SCE). Furthermore, it is evident that bulk charge trapping in flash-annealed high-κ gate dielectrics is not an issue. However, the performance of flash-annealed devices is degraded as compared with that of spike-annealed devices. The likely cause of degraded performance in the flash-annealed devices is identified to be higher interface state density due to thermal stress-induced defects and high-κ induced oxygen vacancies. A post-metallization annealing process is shown to be effective for recovering the mobility loss.

The reduction of Schottky barrier height (SBH) of the silicide to *n*-type silicon contact has been identified as the major challenge to keep pace with contact resistivity projections. In this work, readily available elements such as nitrogen and fluorine are investigated for the passivation of silicide-silicon interface defects for a simpler, low-cost highly manufacturable integration route to achieve low contact resistance technology. NiSi/*n*-type lightly doped Si Schottly diodes were fabricated to study barrier height modulation using nitrogen and fluorine implantation prior to silicidation. It is shown that significant SBH lowering (to 0.3 eV) is achieved on *n*-type silicon using nitrogen. The mechanism of this lowering is explained by interface passivation at lower nitrogen doses and bandgap reduction due to strain at higher nitrogen doses. Effective SBH modulation with nitrogen is a promising method for reduced contact resistance in advanced CMOS technology and provides a solution to achieve $\rho_c < 10^{-8}$ Ω-cm^2 for state-of-the-art silicide-to-source/drain contacts as shown in Figure 6.2. However, very small change in SBH is seen with fluorine implantation which may be due to higher bonding enthalpy of Si-F.

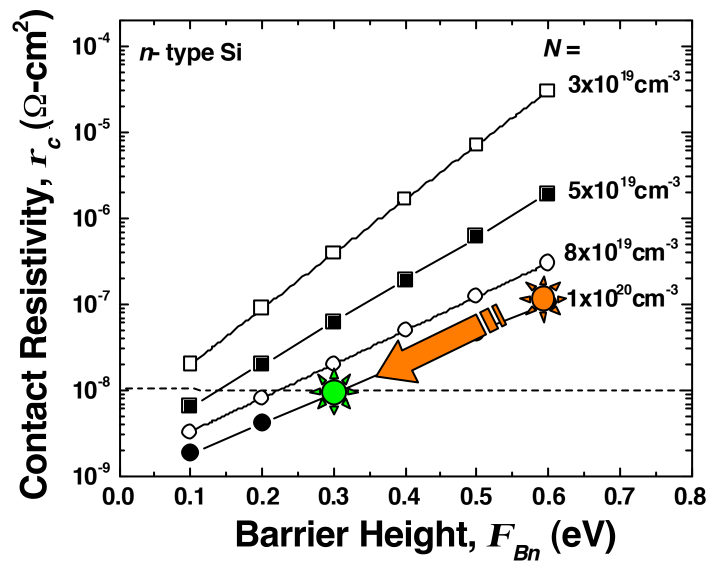


Figure 6.2. With the help of nitrogen implantation prior to silicidation, contact resistivity below 10^{-8} Ω-cm^2 can be obtained for state-of-the-art silicide contacts to heavily doped source/drain.

It appears that the Si channel itself is running out of steam in being able to provide for high-performance needs in highly scaled devices beyond the 22 nm technology generation. Due to higher bulk mobility and demonstrated capability with high-κ dielectrics, epi-Ge PMOSFETs integrated on Si platform provide an exciting option to silicon technology. The material properties of nickel based contact material formed on epi-Ge on Si substrate are investigated and it is shown that low-resistance nickel monogermanide can be formed with very low thermal budget. A wet chemical treatment using sulfuric acid is developed to achieve the self-aligned nickel germanide process. It is demonstrated that the commonly used spacer material such as silicon nitride and TEOS are immune to the developed wet treatment for selective removal of unreacted nickel.

6.3 Recommendations for Future Research

The efficacy of flash annealing to annihilate implantation damage is a primary concern because of the ultra-short annealing time. Furthermore, minimum gate-to-S/D overlap (to avoid hot-carrier reliability issues) is difficult to attain because of diffusion-less nature of the anneal. These two concerns are major roadblocks for the adoption of flash anneal. Therefore, to counter these problems, a low temperature anneal will be needed in addition to flash anneal to provide sufficient thermal budget to satisfy overlap requirement and to annihilate implantation damage to reduce the junction leakage. However, careful optimization of anneal process parameters is required so that dopant deactivation and diffusion don't offset the USJs requirements. Along these lines, flexibly shaped flash annealing has been proposed which provides activation and recovery anneals of the order of milliseconds duration [1].

In this work, the use for flash annealing for low resistance contact technology is not considered. The combination of nitrogen implantation for achieving lower SBH at silicide-source/drain contacts and flash annealing for higher active dopant concentration in source/drain regions can further lower contact resistivity. It is presented that strain-induced barrier lowering and interface passivation are two main mechanisms taking place for SBH lowering due to nitrogen implantation. Image-force induced barrier lowering (using interfacial segregation of dopants) can be further combined with nitrogen implantation and flash anneal to achieve ultra-low resistance contact technology.

Finally, looking into the process development of future CMOS technology generations the approaches mentioned above for SBH lowering (strain, interface pasivation, image force) can be extended to SiGe-based source/drain (for PMOSFETs) and Si:C-based source/drain (for NMOSFETs) to solve the contact resistance problem.

6.4 References

[1] T. Onizawa, S. Kato, T. Aoyama, Y. Nara, and Y. Ohji, "A Proposal of New Concept Milli-second Annealing: Flexibly-Shaped-Pulse Flash Lamp Annealing (FSP-FLA) for Fabrication of Ultra Shallow Junction with Improvement of Metal gate High-k CMOS Performance," *VLSI Symp. Tech. Digest*, pp. 110-111, 2008.