Oxygen-insertion Technology for CMOS Performance Enhancement

Xi Zhang

Electrical Engineering and Computer Sciences
University of California at Berkeley

Technical Report No. UCB/EECS-2019-177
http://www2.eecs.berkeley.edu/Pubs/TechRpts/2019/EECS-2019-177.html

December 18, 2019

Oxygen-insertion Technology for CMOS Performance Enhancement

by

Xi Zhang

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 Vivek Subramanian
Professor Junqiao Wu

Fall 2019

Oxygen-insertion Technology for CMOS Performance Enhancement

Abstract

Oxygen-insertion Technology for CMOS Performance Enhancement

by

Xi Zhang

Doctor of Philosophy in Engineering - Electrical Engineering and Computer Sciences

University of California, Berkeley

Professor Tsu-Jae King Liu, Chair

Until 2003, the semiconductor industry followed Dennard scaling rules to improve complementary metal-oxide-semiconductor (CMOS) transistor performance. However, performance gains with further reductions in transistor gate length are limited by physical effects that do not scale commensurately with device dimensions: short-channel effects (SCE) due to gate-leakage-limited gate-oxide thickness scaling, channel mobility degradation due to enhanced vertical electric fields, increased parasitic resistances due to reductions in source/drain (S/D) contact area, and increased variability in transistor performance due to random dopant fluctuation (RDF) effects and gate work function variations (WFV). These emerging scaling issues, together with increased process complexity and cost, pose severe challenges to maintaining the exponential scaling of transistor dimensions. This dissertation discusses the benefits of oxygen-insertion (OI) technology, a CMOS performance booster, for overcoming these challenges.

The benefit of OI technology to mitigate the increase in sheet resistance (R_{sh}) with decreasing junction depth (X_J) for ultra-shallow-junctions (USJs) relevant for deep-sub-micron planar CMOS transistors is assessed through the fabrication of R_{sh} test structures, electrical characterization, and technology computer-aided design (TCAD) simulations. Experimental and secondary ion mass spectroscopy (SIMS) analyses indicate that OI technology can facilitate low-resistivity USJ formation by reducing R_{sh} and X_J, due to retarded transient-enhanced-diffusion (TED) effects and enhanced dopant retention during post-implantation thermal annealing. It is also shown that a low-temperature-oxide (LTO) capping can increase R_{sh} unfavorably due to lower dopant activation levels, which can be alleviated by OI technology.

This dissertation extends the evaluation of OI technology to advanced FinFET technology, targeting 7/8-nm low power technology node. A bulk-Si FinFET design comprising a super-steep retrograde (SSR) fin channel doping profile achievable with OI technology is studied by three-dimensional (3-D) TCAD simulations. As compared with the conventional bulk-Si (control) FinFET design with a heavily-doped fin channel doping profile, SSR FinFETs can achieve higher I_{on}/I_{off} ratios and reduce the sensitivity of device performance to

variations due to the lightly doped fin channel. As compared with the SOI FinFET design, SSR FinFETs can achieve similarly low $V_{DD,min}$ for 6T-SRAM cell yield estimation. Both SSR and SOI design can provide for as much as 100 mV reduction in $V_{DD,min}$ compared with the control FinFET design. Overall, the SSR FinFET design that can be achieved with OI technology is demonstrated to be a cheaper alternative to the SOI FinFET technology for extending CMOS scaling beyond the 10-nm node.

Finally, this dissertation investigates the benefits of OI technology for reducing the Schottky barrier height (Φ_{Bp}) of a Pt/Ti/p-type Si metal-semiconductor (M/S) contact, which can be expected to help reduce the specific contact resistivity for a p-type silicon contact. Electrical measurements of back-to-back Schottky diodes, SIMS, and X-ray photoelectron spectroscopy (XPS) show that the reduction in Φ_{Bp} is associated with enhanced Ti 2p and Si 2p core energy level shifts. OI technology is shown to favor low-Φ_{Bp} Pt monosilicide formation during forming gas anneal (FGA) by suppressing the grain boundary enhanced diffusion of Pt atoms into the crystalline Si substrate.

To Delin Zhang, Yuling Lai and Yashu Li

for their unbounded love and ceaseless support

Contents

Acknowledgments

I am very grateful to so many people for their help and support, along with my pursuit of a doctoral degree. This dissertation would not have been possible without their help.

First and foremost, I would like to express my sincere gratitude to my advisor Professor Tsu-Jae King Liu, for her inspiring guidance and constant encouragement throughout my graduate program. 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 also like to thank Professor Vivek Subramanian, Professor Junqiao Wu, and Professor Ana Claudia Arias for serving on my dissertation committee and qualifying examination committee to provide valuable advice on my research.

Special thanks go to my research partners with regard to this thesis work. With Hideki Takeuchi, Daniel Connelly, Robert Mears, Marek Hytha, and Rubin Leonard, I could perform my research more elaborately. They provided the kindest assistance and countless discussions on advanced FinFET design and semiconductor process issues. In particular, I would like to thank Daniel Connelly for being an excellent mentor, whose keen perspective on semiconductor TCAD tools and sincere advice were exceptionally helpful. I must also thank Donovan Lee and Carrie Low for being my mentors during my summer internship at Intermolecular in 2016. Their industrial experiences and rigorous technical knowledge have provided a great learning experience. I especially thank Donovan for being a big brother and giving me numerous help, on both research and life.

This work would not have been possible without the efforts of staff members of the UC Berkeley Marvell Nanofabrication Laboratory, ensuring the successful operations and maintenance of equipments, and providing technical assistance and advice. I thank especially Dr. William Flounders, Richelieu Hemphill, Ryan River, Anthony Bielecki, Joanna Bettinger, Allison Dove, and David Lo.

I am sincerely grateful to the excellent members of King Group, the entire Device Group, and the Berkeley EECS department: students, staff, postdoctoral, and visiting scholars. I treasure the interaction we had as colleagues, as well as friends, as our paths crossed at various points of my graduate school experience. Some of them helped me professionally or personally, through action or advice. Some I have learned a thing or two from. Others are simply a joy to interact with and had made my days brighter. I thank especially Prof. Sung Hwan Kim, Prof. Kimihiko Kato, Dr. Zach Jacobson, Dr. Reinaldo Vega, Dr. Nuo Xu, Dr. Chuang Qian, Dr. Peng Zheng, Dr. Ruonan Liu, Dr. Jalal Naghsh Nilchi, Dr. Wei-Chang Li, Dr. Robert Schneider, Dr. Henry Barrow, Dr. Lingqi Wu, Dr. Yi-Ting (Dennis) Wu, Fei Ding, Urmita Sikder, Thomas Rembert, Benjamin Osoba, Alice Ye, Xiaoer Hu, Tsegereda Esatu, Dr. Ruoxi Jia, Dr. Min Jin, Mahsa Sadeghi, Kristel Deems, Peter Zhao, Nishita Deka, Juan Pablo, and Carlos Biaou.

In addition to the above, several others have provided administrative and personal assistance in some form. I am grateful to Nicole Song, and Shirley Salanio, who have been great sources

of information and provided me with various support in fulfilling graduation requirements and difficult situations.
Last but not least, I am tremendously grateful to my family for their unconditional love and support. I am highly indebted to my parents, Delin Zhang and Yuling Lai, who continue to make sacrifices for me, shape and lead me into the individual I am today. Their guidance, support, and endless care have continued to propel me forward through life's numerous challenges. I am exceptionally thankful to my wife, Yashu Li, who with her compassion and understanding helped me through the darkest times, even when she was far away. I am truly thankful for my wonderful family!

Chapter 1

Introduction

1.1 Transistor Scaling Trend

The transistor is arguably the most important invention of the 20th century as it has changed the course of history as the building block for all modern electronics. On Dec. 16th, 1947, John Bardeen, William Shockley, and Walter Brattain at the Bell Telephone Laboratories built the very first "point-contact" transistor, consisting of two gold foil contacts contacting a germanium crystal on a metal plate [1, 2]. This device was designed to supersede the energy-consuming and bulky vacuum tubes. Its invention together with the development of integrated circuits (IC) in 1958 at Texas Instruments [3] soon ignited a series of technology advancements. In 1965, Gordon Moore made a famous economic prediction that the number of transistors per IC would increase exponentially over time based on his empirical observation that unit cost is falling with increasing complexity, as shown in Figure 1.1 [4]. Figure 1.2 shows that "Moore's law" has been used as the roadmap for semiconductor companies to make faster, smaller, and cheaper transistors for the past five decades.

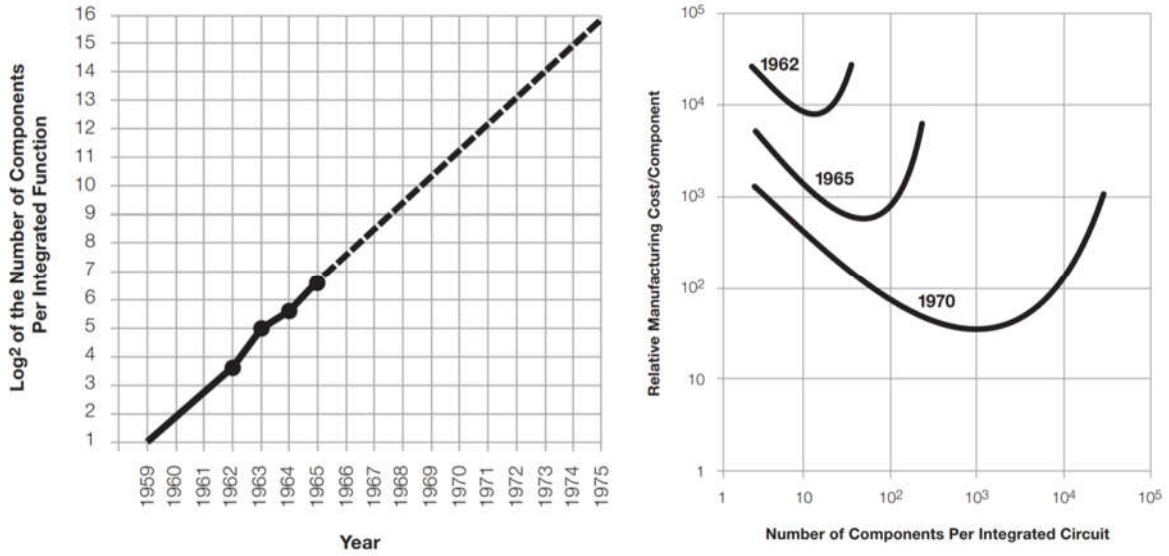

Fig. 1.2. Doubling the number of components per IC is driven by shrinking unit cost (adapted from [4]).

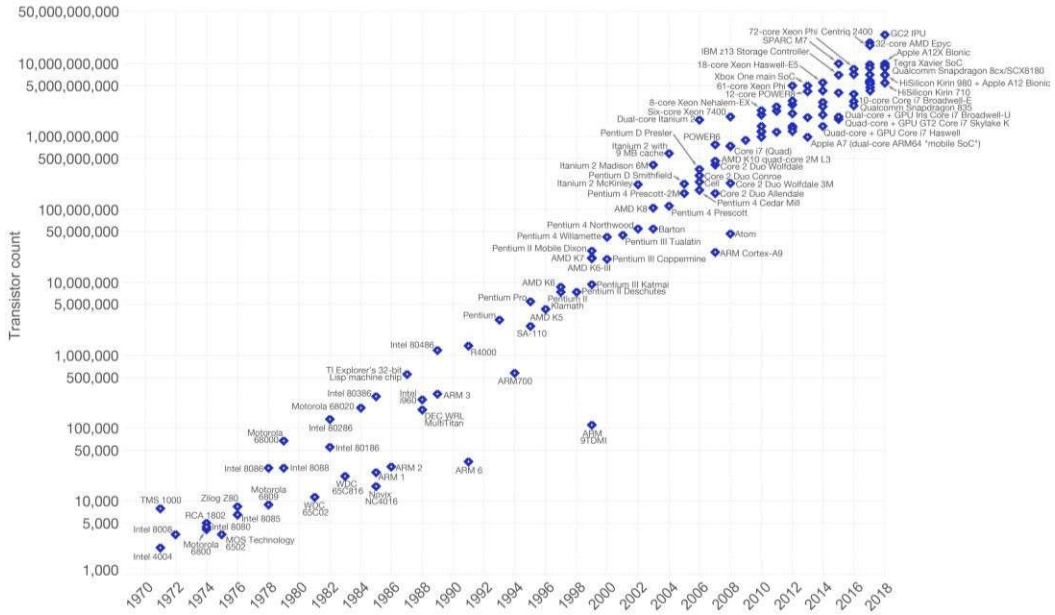

Fig. 1.2. Moore's law: the number of transistors on IC chips from 1971 to 2018 [5].

The exponential scaling of transistors is driven by the exponentially decreasing cost per function, which has slowed down dramatically since 2012, around the time when 22 nm technology node went into volume production. The slowdown is attributed to increased process complexity and cost [6-13], as shown in Figure 1.3.

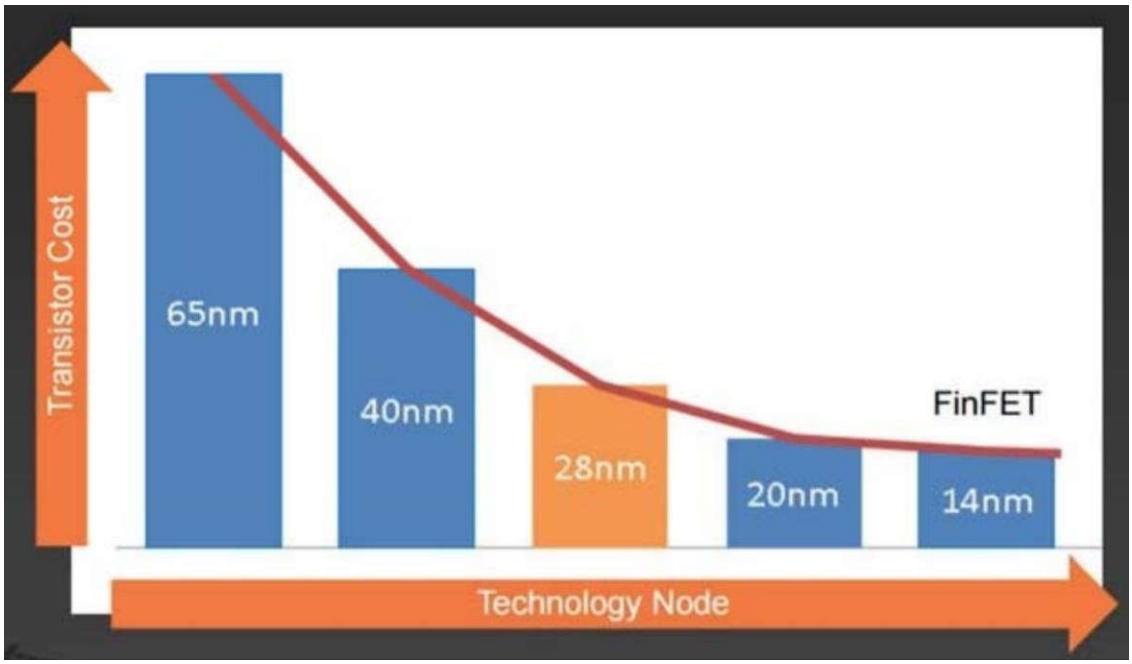

Fig. 1.3. Cost per transistor is no longer decreasing (adapted from [4]).

1.2 Complementary Metal-oxide-semiconductor (CMOS) Technology

Figure 1.4 shows the basic structure of an N-channel metal-oxide-semiconductor field-effect transistor (N-MOSFET, or NFET). In NFETs, source (S) and drain (D) regions are n+ doped by introducing group VI atoms (arsenic, phosphorus) into Si lattice. The Si substrate is doped p-type by incorporating group III (boron) atoms. "N-channel" refers to the conducting n-type (electron-rich) pathway in the ON state: an inversion layer forms in the channel region with gate-to-source voltage (V_{GS}) higher than threshold voltage (V_{th}) and drain-to-source voltage (V_{DS}) $> 0\ V$. V_{th} is defined as the gate-to-source voltage needed to turn on the device.

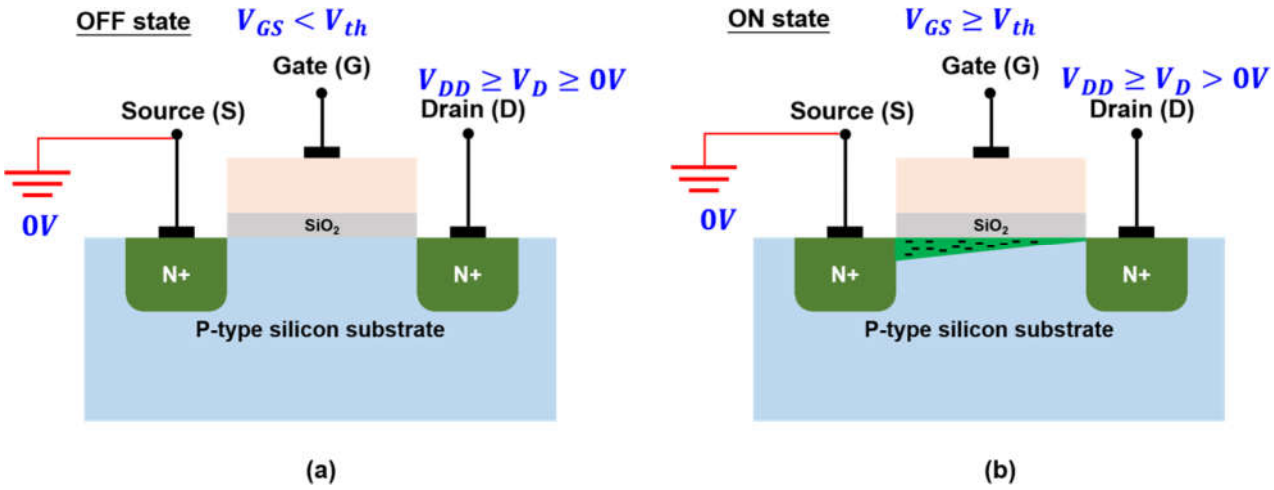

Fig. 1.4. (a) Cross-sectional schematic drawing of an NFET in the OFF state ($V_{GS} < V_{th}$, $V_{DS} \geq 0V$); (b) NFET in the ON state ($V_{GS} \geq V_{th}$, $V_{DS} > 0V$).

Figure 1.5 illustrates a P-channel MOSFET (PFET) with p+ S/D regions and an n-type Si substrate. "P-channel" indicates the formation of a p-type (hole-rich) current conduction pathway in its ON state. In both NFET and PFET, drain voltage (V_D) and gate voltage (V_G) swing between 0V and V_{DD}, with the source voltage (V_S) tied to the ground (0V) and the power supply (V_{DD}) rails, respectively. Therefore, in static states, the PN junctions between S/D regions and the substrate are always reverse-biased and do not conduct forward diode current. Essentially, MOSFET can be viewed as a voltage-controlled switch: conducting high current (I_{on}) when turned ON and minimal leakage current (I_{off}) when turned OFF. Figure 1.6 shows MOSFET basic current-voltage (IV) characteristics.

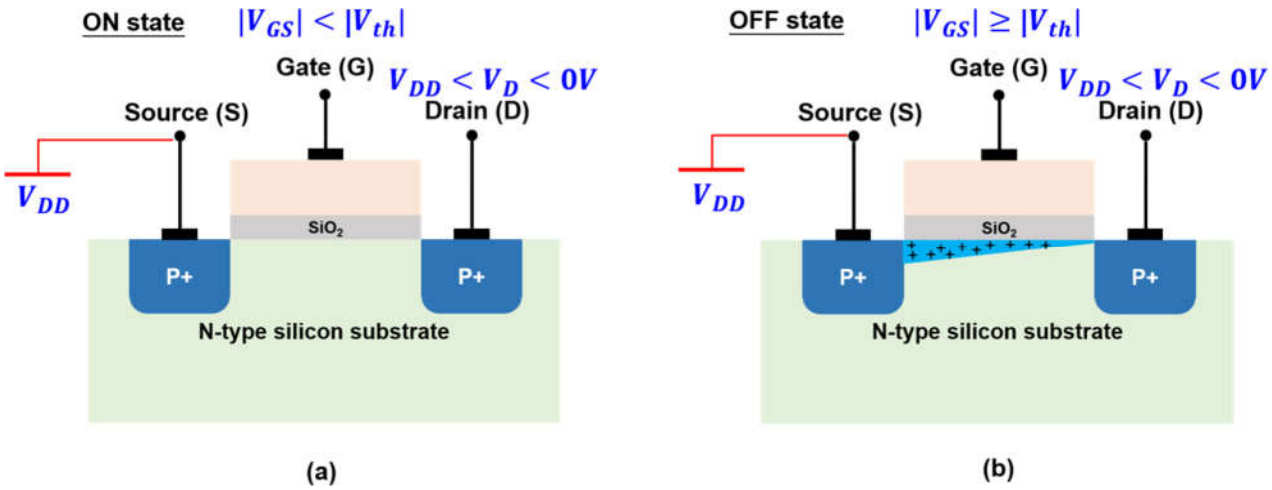

Fig. 1.5. (a) Cross-sectional schematic drawing of an PFET in the OFF state ($|V_{GS}| < |V_{th}|$, $|V_{DS}| \geq 0V$); (b) PFET in the ON state ($|V_{GS}| \geq |V_{th}|$, $|V_{DS}| > 0V$).

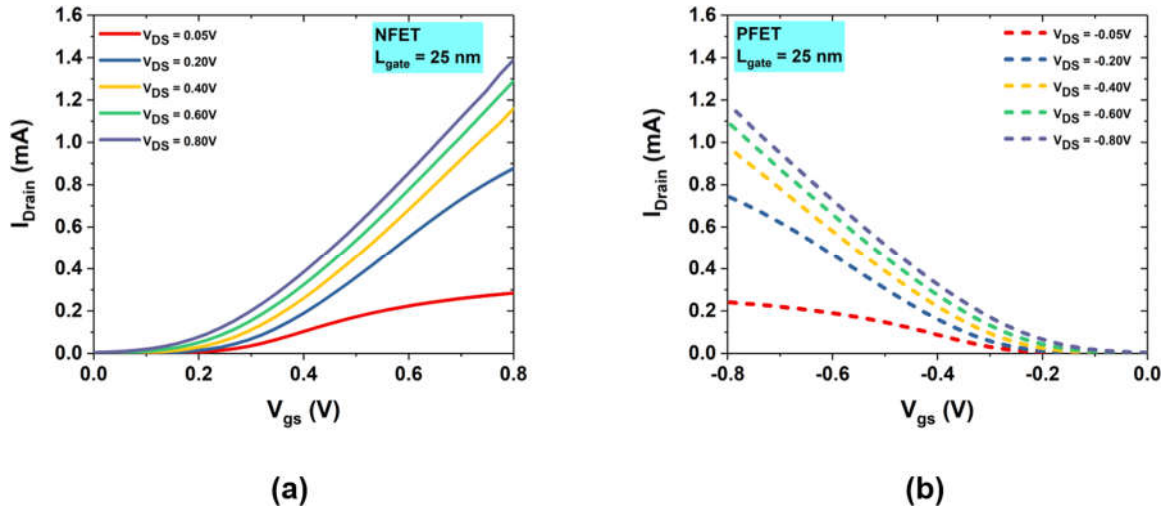

Fig. 1.6. Simulated transfer IV characteristics (drain current I_{Drain} vs. gate-to-source voltage V_{GS}) of (a) NFET (b) PFET.

The complementary operation of n-channel and p-channel MOSFETs depicted in Figure 1.6 allows engineers to design complementary MOS (CMOS) circuits with low static power consumption. Figure 1.7 demonstrates a CMOS inverter, one of the most basic building blocks in modern digital IC circuits, along with the corresponding voltage transfer characteristic (VTC).

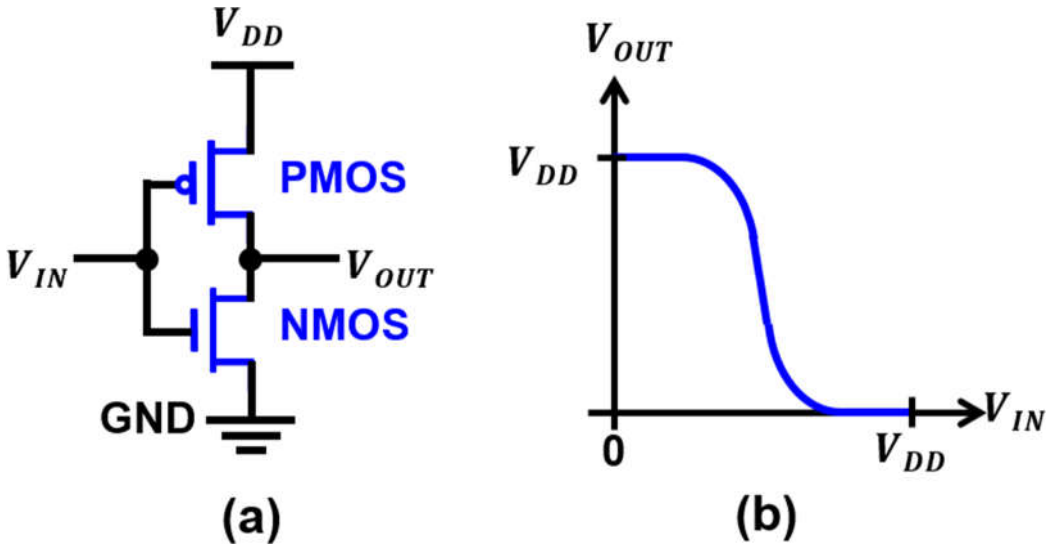

Fig. 1.7. (a) CMOS inverter, V_{DD} stands for the power supply voltage; (b) typical VTC of a CMOS inverter.

When the input voltage is high ($V_{IN} = V_{DD}$), the NMOS transistor is turned ON, and the PMOS transistor is turned OFF. This results in a current pathway between the output node (V_{OUT}) and the ground, discharging the output node to 0V. Similarly, when V_{IN} is 0V, the PMOS is turned ON, and the NMOS is turned OFF, charging V_{OUT} to V_{DD}. Under static conditions, no direct current pathway (except the small OFF-state leakage current path) exists between the power supply and ground rails, hence mitigating static power consumption. Hence, the switching power consumption (charging and discharging circuit capacitances) contributes most significantly to total power consumption. In general, total power consumption (P_{total}) for CMOS circuits can be divided into two categories: dynamic power consumption ($P_{dynamic}$) and static power consumption (P_{static}). The dependence of P_{total} on circuit parameters is given by the following equation:

$$\begin{aligned} P_{total} &= P_{static} + P_{dynamic} = P_{static} + P_{switch} + P_{sc} \\ &= I_{OFF}V_{DD} + \alpha f C V_{DD}^2 + I_{SC}V_{DD}t_{SC}f \end{aligned} \tag{1.1}$$

Where I_{OFF} is the off-state leakage current, α is the switching "activity factor" (defined as the average percentage of circuit capacitance switched per clock cycle), C is the total circuit capacitance, f is the clock frequency, I_{sc} is the short-circuit current spike when pull-up (PMOS) and pull-down (NMOS) networks are ON simultaneously due to the non-abrupt switching slope of input voltage signal, and t_{sc} is the average time duration of I_{sc} per clock cycle.

Another important figure of merit to gauge how fast a CMOS circuit operates is the propagation delay (t_p). It is directly related to the product of circuit capacitance (C) and

transistor ON-state resistance (R_{ON}). The latter is inversely proportional to the channel width to length (W/L) ratio for planar MOSFETs or the number of fins for FinFETs:

$$t_P \propto R_{ON} C \tag{1.2}$$

To build faster circuits with lower power dissipation, it is desirable to reduce t_p and P_{total} simultaneously.

CMOS power dissipation can be minimized in the following ways: (1) reducing V_{DD}, which has a quadratic influence on $P_{dynamic}$; (2) reducing f, which slows down the circuit operation undesirably; (3) lowering C, which motivates the continued scaling of transistor dimensions as transistor capacitances contribute significantly to C; (4) reducing α, which is accomplished at the logic circuit and architecture-abstraction levels, such as: slowing down non-critical path, power gating, parallelism and pipelining.

To minimize the propagation delay, possible solutions are: (1) reducing circuit capacitances (drain diffusion capacitances, interconnect capacitances, and fan-out); (2) increasing I_{ON}, which can be achieved by using a larger W/L ratio for planar MOSFETs or increasing the number of fins for FinFETs; (3) increasing V_{DD}, which can reduce charging/discharging time of circuit capacitances by increasing the overdrive voltage ($|V_{ov}| = |V_{gs} - V_{th}|$).

A design trade-off becomes evident from the above analysis. Lower V_{DD} is desirable for lower power dissipation. One can choose a lower threshold voltage (V_{th}) value to compensate for the loss in transistor performance. However, V_{th} is lower-bounded by the requirement to minimize static power dissipation. In other words, choosing V_{th} represents a trade-off between static power dissipation and performance, as depicted in Figure 1.8.

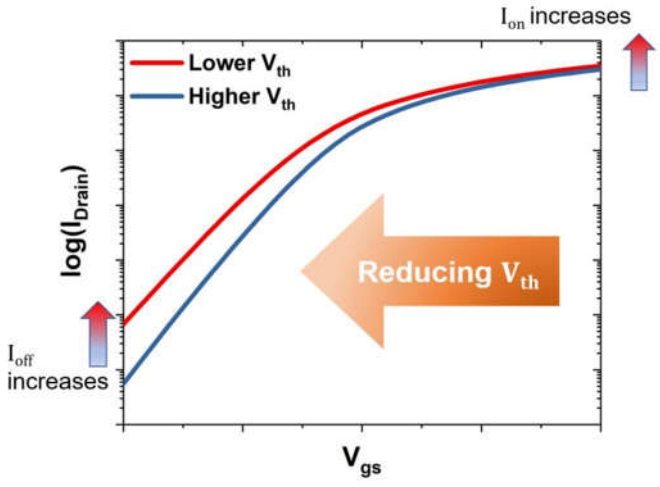

Fig. 1.8. V_{th} design tradeoff: a higher V_{th} value reduces static power dissipation but also slows down circuits; a lower V_{th} value improves circuit speed but results in higher static power dissipation.

Lower circuit capacitances improve both circuit power dissipation and speed, which is one of the major motivations behind the continued miniaturization of MOSFETs.

1.3 CMOS Scaling Challenges

Table 1.1 shows that until 2003, the semiconductor industry has been following Dennard scaling for three decades to improve CMOS performance [14-17]. Voltage reduction per generation is mandatory to avoid significant increases in power density. This indicates the necessity to reduce threshold voltage (V_{th}) in proportion to V_{DD} reduction to maintain the same current over-drive (V_{DD} - V_{th}) for achieving at least the same on-state current (I_{on}). However, V_{th} cannot be scaled arbitrarily small because of power consumption constraints (I_{off} needs to be reasonably low to suppress passive power consumption). As a result, V_{DD} reduction per generation has decreased, and the passive circuit power consumption has now become a dominant contributor to the overall power consumption. This prevents semiconductor companies from following Dennard scaling to improve transistor performance since the 130nm technology node [18, 19].

Parameter	Constant field scaling	Generalized scaling
Physical dimensions: L_{gate}, W, T_{ox}, wire pitch	$1/\alpha$	$1/\alpha$
Body doping concentration	α	E/α
Voltage	$1/\alpha$	E/α
Circuit density	$1/\alpha^2$	$1/\alpha^2$
Capacitance per circuit	$1/\alpha$	$1/\alpha$
Circuit speed	α	α
Circuit power	$1/\alpha^2$	E^2/α^2
Power density	1	E^2
Power-delay product	$1/\alpha^2$	E^2/α^3

Table 1.1. CMOS scaling scenarios. α is the scaling constant, greater than 1, for device dimensions, $E = V/\alpha$ is the normalized electric field (adapted from [17]).

In addition to the power density constraints [18, 20], several other scaling issues have emerged as the physical dimensions of CMOS transistors approach the atomic scale. These limitations are imposed by physical effects that do not scale commensurately with device

dimensions: (1) channel mobility degradation due to enhanced vertical electric fields [6, 7, 21-23]; (2) gate leakage current due to oxide thickness scaling [24-27]; (3) short-channel effects [28-30]; (4) parasitic resistances and capacitances [31-34]; (5) transistor variations such as random dopant fluctuation effects [35-37] and gate work function variations [38-40]. Alternative approaches have been adopted through the years to overcome these challenges and achieve performance improvements. Figure 1.9 illustrates major performance-boosters adopted by Intel Corporation for enhancing carrier mobility, suppressing gate leakage current, and achieving superior electrostatic control [6, 7, 21-23].

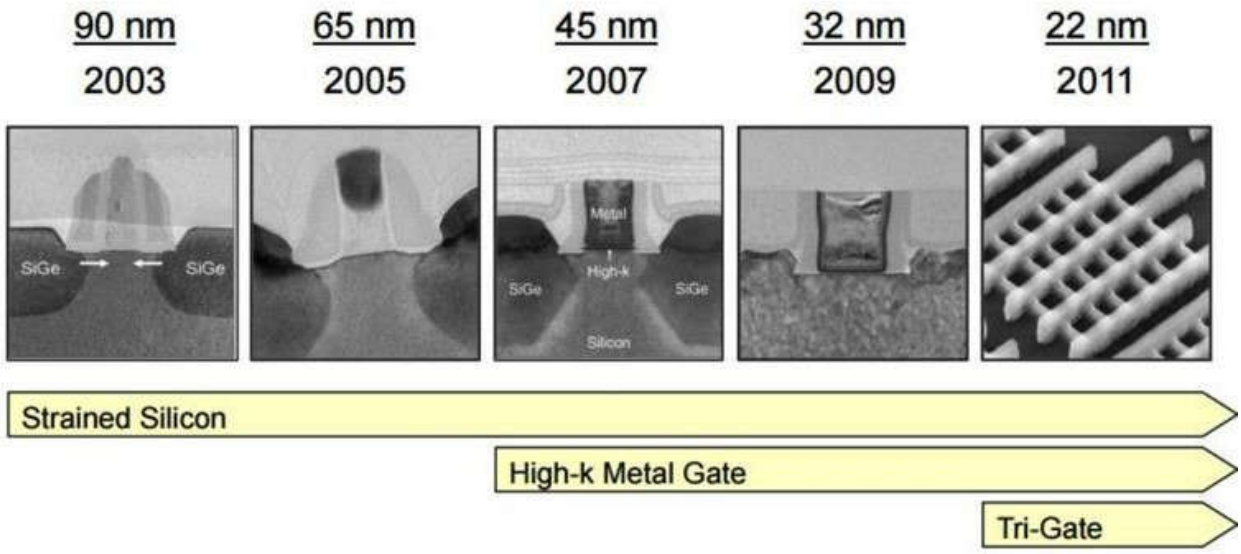

Fig. 1.9. CMOS performance boosters invented in the non-classical scaling era: strained silicon channel [21], high-k dielectric and metal gate materials [22], tri-gate structures [6] (adapted from [41]).

1.4 Research Objectives

This dissertation discusses the application of "oxygen insertion" (OI) technology to facilitate the scaling of CMOS transistors [42]. Its benefits for both advanced planar MOSFET and tri-gate MOSFET (FinFET) technologies are evaluated through experiments and technology computer-aided design (TCAD) simulations.

In Chapter 2, a comprehensive literature review on OI technology is provided, and its physical mechanisms are discussed. Since past research work showed that OI technology is promising to improve MOSFET performance, OI technology is chosen as the focus of this dissertation.

In Chapter 3, the impacts of various capping layers, including OI layers, on ultra-

shallow junction (USJ) formation are studied. First, experimental results and TCAD simulation data are presented to compare the benefits of various capping layers on dopant diffusion and dopant activation. Then, the physical mechanisms behind the improved USJ characteristics due to OI technology are discussed.

In Chapter 4, a TCAD-based simulation study is presented to quantify the benefits of a super-steep retrograde (SSR) fin channel doping profile achievable with OI technology, targeting 7/8-nm low-power applications. Next, the electrostatic benefits of using a silicon-on-insulator (SOI) substrate versus the bulk-silicon SSR FinFET technology are investigated via Sentaurus Device TCAD. Finally, the benefits of SSR and SOI FinFET technologies for 6-transistor static RAM (6T-SRAM) cell performance and yield are estimated using a calibrated compact model.

In Chapter 5, the effects of oxygen-insertion (OI) technology and low-energy fluorine (F) implantation on the Schottky barrier height (Φ_{Bp}) of a Pt/Ti/p-type Si metal-semiconductor (M/S) contact are investigated. First, electrical characterization results of Schottky diodes were presented to compare the benefits of OI layers and F for Φ_{Bp} reduction. Then, X-ray photoelectron spectroscopy (XPS) and secondary ion mass spectroscopy (SIMS) analyses were performed to elucidate the roles of oxygen and fluorine on Ti and Pt diffusion into Si, as well as on Pt silicidation.

In Chapter 6, the contributions of this work are summarized, and the future directions for further research are suggested.

1.5 References

[1] B. John and W. H. Brattain, "Three-electrode circuit element utilizing semiconductive materials," ed: Google Patents, 1950.

[2] J. Bardeen and W. H. Brattain, "The transistor, a semi-conductor triode," *Physical Review,* vol. 74, no. 2, p. 230, 1948.

[3] J. S. Kilby and J. S. Kilby, "Invention of the integrated circuit," *Resonance,* vol. 17, no. 11, 1976.

[4] G. E. Moore, "Cramming more components onto integrated circuits," ed: McGraw-Hill New York, NY, USA, 1965.

[5] M. Roser and H. Ritchie, "Technological progress," *Our World in Data,* 2013.

[6] C. Auth *et al.*, "A 22nm high performance and low-power CMOS technology featuring fully-depleted tri-gate transistors, self-aligned contacts and high density MIM capacitors," in *2012 Symposium on VLSI Technology (VLSIT)*, 2012: IEEE, pp. 131-132.

[7] S. Narasimha *et al.*, "22nm High-performance SOI technology featuring dual-embedded stressors, Epi-Plate High-K deep-trench embedded DRAM and self-aligned Via 15LM BEOL," in *2012 International Electron Devices Meeting*, 2012: IEEE, pp. 3.3. 1-3.3. 4.

[8] A. E.-J. Lim *et al.*, "Review of silicon photonics foundry efforts," *IEEE Journal of Selected Topics in Quantum Electronics,* vol. 20, no. 4, pp. 405-416, 2013.

[9] K. Flamm, "Measuring Moore's law: Evidence from price, cost, and quality indexes,"

National Bureau of Economic Research, 2018.
[10] K.-Y. Lim *et al.*, "Novel stress-memorization-technology (SMT) for high electron mobility enhancement of gate last high-k/metal gate devices," in *2010 International Electron Devices Meeting*, 2010: IEEE, pp. 10.1. 1-10.1. 4.
[11] N. Xu, B. Ho, M. Choi, V. Moroz, and T.-J. K. Liu, "Effectiveness of stressors in aggressively scaled FinFETs," *IEEE T Electron Dev,* vol. 59, no. 6, pp. 1592-1598, 2012.
[12] S. Eneman *et al.*, "Layout impact on the performance of a locallystrained PMOSFET," in *Digest of Technical Papers. 2005 Symposium on VLSI Technology, 2005.*, 2005: IEEE, pp. 22-23.
[13] N. Xu, L. T.-N. Wang, A. R. Neureuther, and T.-J. K. Liu, "Physically based modeling of stress-induced variation in nanoscale transistor performance," *IEEE Transactions on Device and Materials Reliability,* vol. 11, no. 3, pp. 378-386, 2011.
[14] G. Baccarani, M. R. Wordeman, and R. H. Dennard, "Generalized scaling theory and its application to a ¼ micrometer MOSFET design," *IEEE T Electron Dev,* vol. 31, no. 4, pp. 452-462, 1984.
[15] B. Davari, R. H. Dennard, and G. G. Shahidi, "CMOS scaling for high performance and low power-the next ten years," *Proceedings of the IEEE,* vol. 83, no. 4, pp. 595-606, 1995, doi: 10.1109/5.371968.
[16] R. H. Dennard, F. H. Gaensslen, V. L. Rideout, E. Bassous, and A. R. LeBlanc, "Design of ion-implanted MOSFET's with very small physical dimensions - IEEE Journals & Magazine," *IEEE Journal of Solid-State Circuits,* vol. 9, no. 5, pp. 256-268, 1974, doi: 10.1109/JSSC.1974.1050511.
[17] W. Haensch *et al.*, "Silicon CMOS devices beyond scaling," *IBM Journal of Research and Development,* vol. 50, no. 4.5, pp. 339-361, 2006.
[18] M. Bohr, "A 30 year retrospective on Dennard's MOSFET scaling paper," *IEEE Solid-State Circuits Society Newsletter,* vol. 12, no. 1, pp. 11-13, 2007.
[19] M. Bohr, "The new era of scaling in an SoC world," in *2009 IEEE International Solid-State Circuits Conference-Digest of Technical Papers*, 2009: IEEE, pp. 23-28.
[20] D. J. Frank, "Power-constrained CMOS scaling limits," *IBM Journal of Research and Development,* vol. 46, no. 2.3, pp. 235-244, 2002.
[21] S. Thompson *et al.*, "A 90 nm logic technology featuring 50 nm strained silicon channel transistors, 7 layers of Cu interconnects, low k ILD, and 1 μm^2 SRAM cell," in *Digest. International Electron Devices Meeting*, 2002: IEEE, pp. 61-64.
[22] K. Mistry *et al.*, "A 45nm logic technology with high-k+ metal gate transistors, strained silicon, 9 Cu interconnect layers, 193nm dry patterning, and 100% Pb-free packaging," in *2007 IEEE International Electron Devices Meeting*, 2007: IEEE, pp. 247-250.
[23] C. Bencher, Y. Chen, H. Dai, W. Montgomery, and L. Huli, "22nm half-pitch patterning by CVD spacer self alignment double patterning (SADP)," in *Optical Microlithography XXI*, 2008, vol. 6924: International Society for Optics and Photonics, p. 69244E.
[24] S. Thompson, "MOS scaling: Transistor challenges for the 21st century," in *Intel Technology Journal*, 1998.
[25] R. Chau *et al.*, "Application of high-κ gate dielectrics and metal gate electrodes to enable silicon and non-silicon logic nanotechnology," *Microelectronic Engineering,* vol. 80, pp. 1-6, 2005.
[26] Y.-C. Yeo, T.-J. King, and C. Hu, "Direct tunneling leakage current and scalability of

alternative gate dielectrics," *Applied Physics Letters,* vol. 81, no. 11, pp. 2091-2093, 2002.

[27] T. Ghani *et al.*, "Scaling challenges and device design requirements for high performance sub-50 nm gate length planar CMOS transistors," in *Symposium on VLSI Technology*, Honolulu, HI, USA, 13-15 June 2000 2000: IEEE, doi: 10.1109/VLSIT.2000.852814.

[28] S.-H. Oh, D. Monroe, and J. Hergenrother, "Analytic description of short-channel effects in fully-depleted double-gate and cylindrical, surrounding-gate MOSFETs," *IEEE electron device letters,* vol. 21, no. 9, pp. 445-447, 2000.

[29] D. J. Frank, R. H. Dennard, E. Nowak, P. M. Solomon, Y. Taur, and H.-S. P. Wong, "Device scaling limits of Si MOSFETs and their application dependencies," *Proceedings of the IEEE,* vol. 89, no. 3, pp. 259-288, 2001.

[30] A. Chaudhry and M. J. Kumar, "Controlling short-channel effects in deep-submicron SOI MOSFETs for improved reliability: a review," *IEEE Transactions on Device and Materials Reliability,* vol. 4, no. 1, pp. 99-109, 2004.

[31] K. K. Ng and W. Lynch, "The impact of intrinsic series resistance on MOSFET scaling," *IEEE T Electron Dev,* vol. 34, no. 3, pp. 503-511, 1987.

[32] B. Doyle *et al.*, "Tri-gate fully-depleted CMOS transistors: Fabrication, design and layout," in *2003 Symposium on VLSI Technology. Digest of Technical Papers (IEEE Cat. No. 03CH37407)*, 2003: IEEE, pp. 133-134.

[33] A. V. Thean *et al.*, "Performance and variability comparisons between multi-gate FETs and planar SOI transistors," in *2006 International Electron Devices Meeting*, 2006: IEEE, pp. 1-4.

[34] A. Khakifirooz and D. A. Antoniadis, "MOSFET performance scaling—Part II: Future directions," *IEEE T Electron Dev,* vol. 55, no. 6, pp. 1401-1408, 2008.

[35] A. Asenov, "Random dopant induced threshold voltage lowering and fluctuations in sub-0.1 μm MOSFET's: A 3-D "atomistic" simulation study," *IEEE T Electron Dev,* vol. 45, no. 12, pp. 2505-2513, 1998, doi: 10.1109/16.735728.

[36] A. Wettstein, O. Penzin, E. Lyumkis, and W. Fichtner, "Random dopant fluctuation modelling with the impedance field method," in *International Conference on Simulation of Semiconductor Processes and Devices, 2003. SISPAD 2003.*, 2003: IEEE, pp. 91-94.

[37] X. Wang, A. R. Brown, B. Cheng, and A. Asenov, "Statistical variability and reliability in nanoscale FinFETs," in *2011 International Electron Devices Meeting*, 2011: IEEE, pp. 5.4. 1-5.4. 4.

[38] K. J. Kuhn, "Reducing variation in advanced logic technologies: Approaches to process and design for manufacturability of nanoscale CMOS," in *2007 IEEE International Electron Devices Meeting*, 2007: IEEE, pp. 471-474.

[39] X. Zhang *et al.*, "Physical model of the impact of metal grain work function variability on emerging dual metal gate MOSFETs and its implication for SRAM reliability," in *2009 IEEE International Electron Devices Meeting (IEDM)*, 2009: IEEE, pp. 1-4.

[40] H. F. Dadgour, K. Endo, V. K. De, and K. Banerjee, "Grain-orientation induced work function variation in nanoscale metal-gate transistors—Part I: Modeling, analysis, and experimental validation," *IEEE T Electron Dev,* vol. 57, no. 10, pp. 2504-2514, 2010.

[41] S. Thompson, "Advance device concepts for 7nm node and beyond," in *Tutorial of International Electron Devices Meeting*, 2015, vol. 2015.
[42] R. J. Mears *et al.*, "Silicon superlattice on SOI for high mobility and reduced leakage," in *2007 IEEE International SOI Conference*, 2007: IEEE, pp. 23-24.

Chapter 2

Oxygen Insertion Technology

In this chapter, oxygen insertion (OI) technology is introduced; then, the physical mechanisms of OI technology to enhance transistor performance via carrier sub-band engineering and dopant profile engineering are discussed.

2.1 Introduction

Historically, simply scaling the transistor dimensions has been adequate to provide for stable integrated circuit (IC) performance improvement while reducing the cost of manufacturing each transistor (cost per function). However, as gate length (L_{gate}) was scaled to below 90 nm, the constant field scaling approach proposed by Robert Dennard [1, 2] slowed down dramatically due to worsening short-channel effects (SCE) [3-6] and transistor variability [7-12]. To overcome these challenges, multiple transistor performance boosters have been adopted over the years, such as doping profile engineering to improve electrostatic integrity and reduce variability [13, 14], strain engineering to enhance channel mobility [15, 16], and high-k/metal-gate (HKMG) materials to enhance gate coupling without increasing gate leakage [17, 18]. OI technology has been proposed to provide for multiple benefits simultaneously, including channel mobility enhancement, gate leakage reduction, electrostatic integrity improvement, and variability reduction [19-31].

In OI technology, multiple partial oxygen monolayers are incorporated into silicon substrates using a low-temperature ($\leq$ 800℃) chemical-vapor deposition (CVD) epitaxy process. To meet semiconductor manufacturing requirements, over 1000 wafers were processed to characterize the associated process control metrology completely [26]. Figure 2.1 shows the depth profiling data for a wafer with four partial oxygen monolayers obtained by secondary ion mass spectroscopy (SIMS).

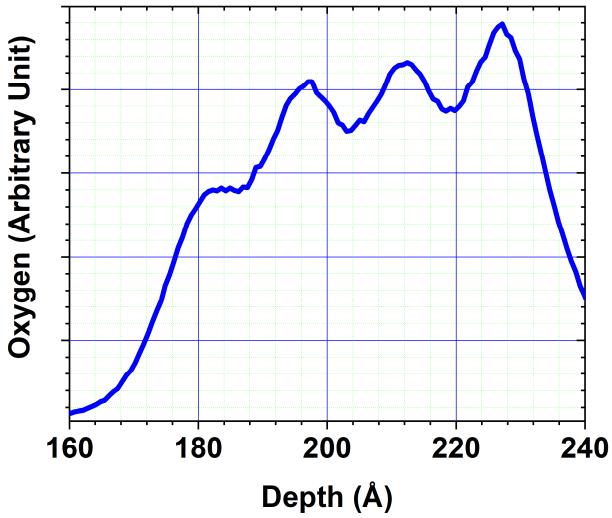

Fig. 2.1. Concentration depth profiles of oxygen atoms (linear scale) obtained by SIMS. Four partial oxygen monolayers are incorporated into the crystalline Si wafer sample.

Within each partial oxygen monolayer, oxygen atoms are incorporated interstitially to guarantee negligible disturbance to the crystalline silicon lattice, allowing subsequent silicon epitaxial growth. Previous research shows OI technology is beneficial for improving carrier mobility and electrostatic integrity. The OI layers are also verified to reduce dopant diffusion and hence help form retrograde channel doping profiles, improving carrier mobility and reducing variability for both planar bulk MOSFETs and advanced FinFETs. OI technology is compatible with other performance-boosting technologies and can be easily integrated into the conventional MOSFET process flow.

2.2 Carrier Sub-band Engineering

With transistor dimension scaling, increased process complexity [32, 33] and shrinking embedded source/drain (S/D) stressor volumes [34, 35] limit the channel mobility improvements provided by strain engineering, requiring alternative mobility enhancement approaches. OI technology has been verified experimentally to produce a local "quantum-confinement" effect on carrier wavefunctions, providing for carrier mobility enhancement and gate leakage reduction [23, 26, 27]. Partial oxygen monolayers can be inserted interstitially into the silicon channel region after the shallow trench isolation (STI) step in a typical planar bulk Si MOSFET process flow. OI technology can be expected to help reduce scattering rates due to separated carrier wavefunction distributions, as shown in Figure 2.2. A high-resolution transmission electron microscopy (TEM) image of the

inserted oxygen monolayers is shown in Figure 2.3.

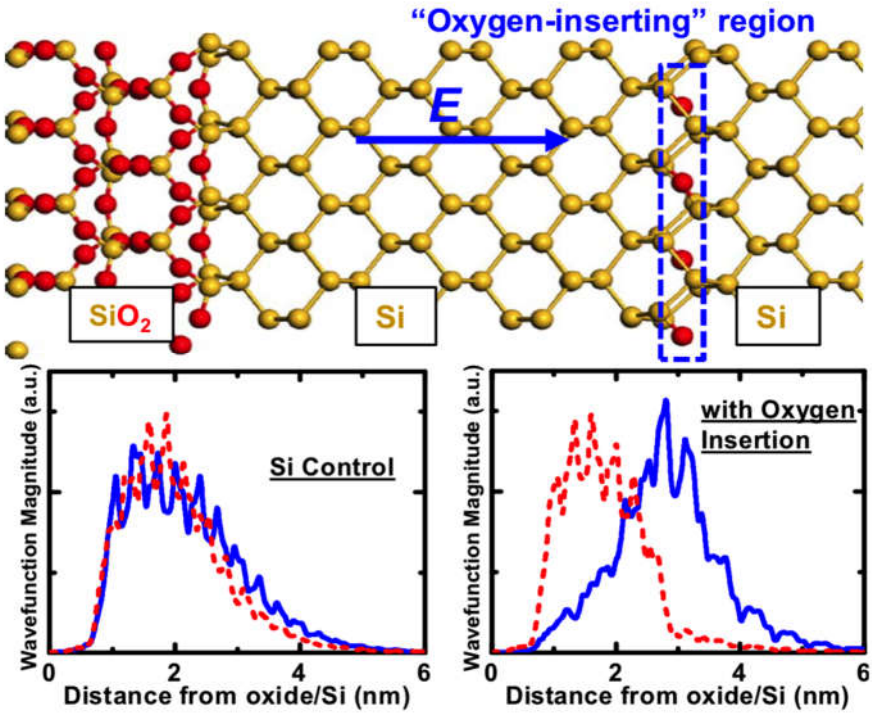

Fig. 2.2. Schematic view of OI layer in Si MOSFET channel region. Ab-initio simulations suggest that carrier wavefunction separation is beneficial for reducing scattering rates [36, 37].

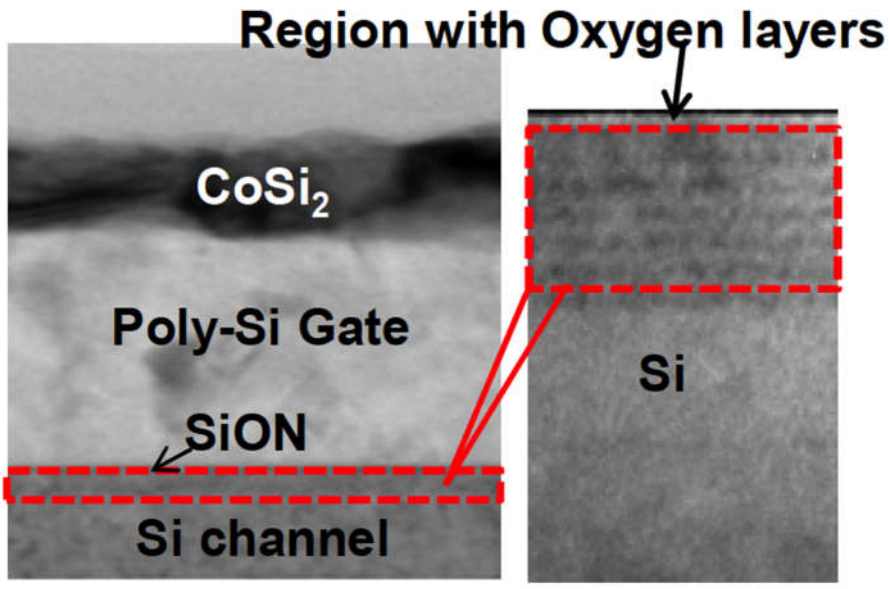

Fig. 2.3. Cross-sectional TEM image of the OI region into planar bulk Si MOSFET channel region. The gate oxide has received plasma nitridation treatment [23].

To investigate the impacts of the OI layers on carrier sub-band structures and carrier mobility, simulations were performed using a state-of-art Poisson-Schrödinger self-consistent simulator [23]. To take into account the impact of the normal/shear stress on band structure and non-parabolic *E-k* effects, the effective mass approximation (EMA) approach is used for electrons [38]. The 6×6 k·p method is adopted for hole simulations [39]. To study phonon and surface roughness scattering mechanisms and dielectric screening effects, field-effect mobility is computed using the Kubo-Greenwood formula [40]. Carrier mobility simulation parameters are calibrated to reproduce the experimental data from [41]. To model electron and hole mobility improvements, inserted partial oxygen monolayers are treated as wide-bandgap energy barriers. The corresponding band gap heights and widths are calibrated to match the measurement results. Figure 2.4 [23] demonstrates this modeling approach can accurately capture the mobility enhancements induced by OI technology.

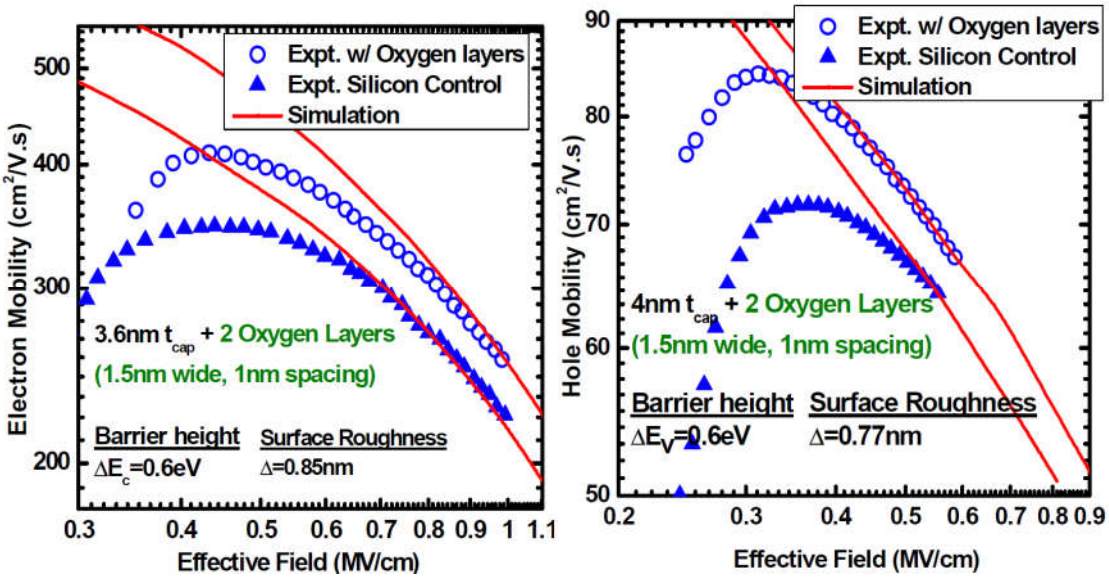

Fig. 2.4. Comparison of simulated carrier universal mobility curves (electron: left; hole: right) vs. experimental data from [41].

Simulated electron/hole sub-band wavefunction magnitudes and conduction/valence band edge energies are plotted against the distance from the top oxide/Si interface in Figure 2.5. OI technology can increase the separation between different sub-band wavefunctions and thus reduce inter-band scattering for both electrons and holes. This can explain the carrier mobility enhancements shown in Figure 2.4 [42].

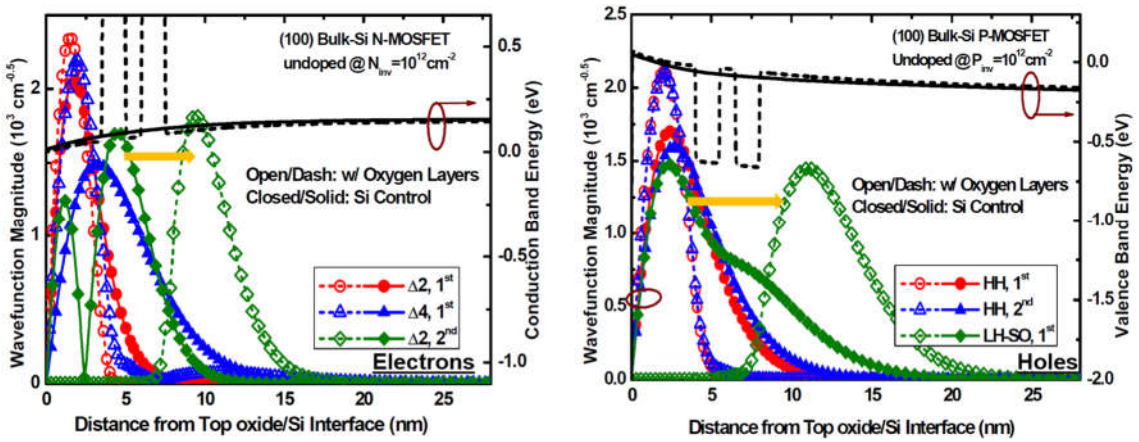

Fig. 2.5. Simulated carrier sub-band's wavefunction magnitudes and band edge energies along the confinement direction, for bulk-Si n-channel (left) and p-channel (right) MOSFETs with and without oxygen layers [23]. OI technology increases the separation between the different $\Delta - 2$ sub-band wavefunctions, reducing phonon form factor and improving electron mobility. Similarly, the separation between heavy-hole (HH) and light hole-split-off (LH-SO) sub-band wavefunctions increases in the presence of the OI layers, reducing the inter-band scattering rate for holes.

Figure 2.6 shows the simulated carrier sub-band energy shifts induced by the OI layers. Further electron mobility enhancement can be expected with OI technology as a result of electrons repopulating to $\Delta - 2$ valleys with lower transport effective mass, which is not observed for holes. The OI layers also help reduce the tunneling effective mass, which is verified by the measured gate leakage currents as seen in Figure 2.7.

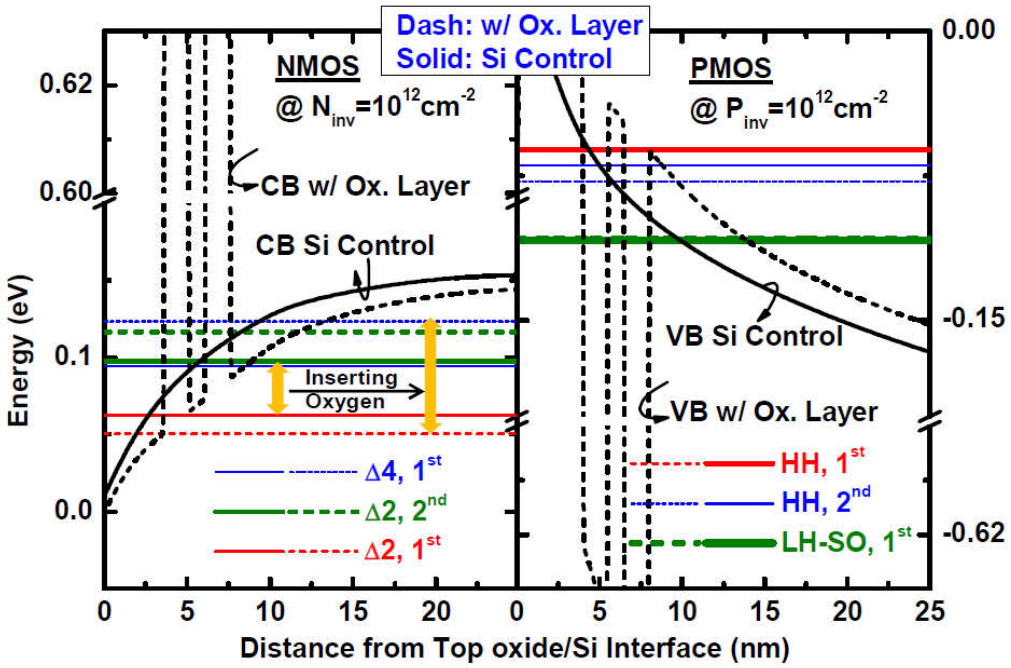

Fig. 2.6. Simulated electron (left) and hole (right) sub-band energies change induced by OI technology.

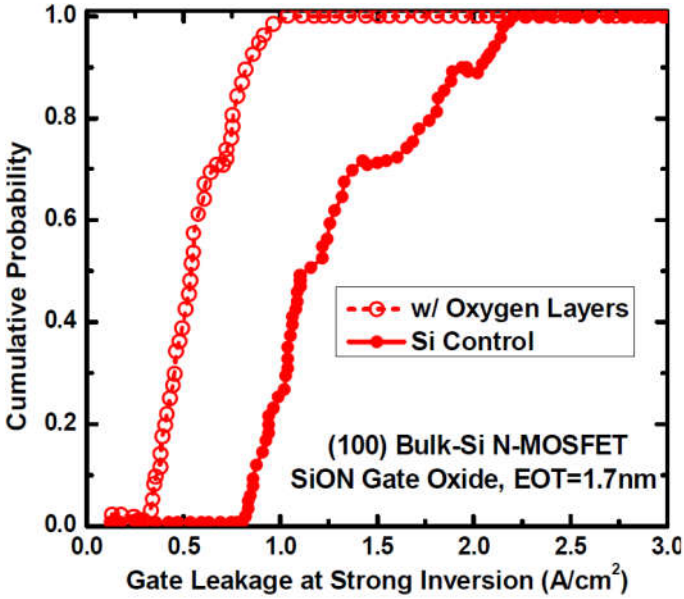

Fig. 2.7. Measured gate leakage cumulative probability plots comparison between bulk-Si N-MOSFETs w/ and w/o oxygen layers.

The main conclusions of the work studying the impact of the inserted oxygen layers on carrier sub-band distributions and mobilities [23] can be summarized as follows.

(1) Insertion of the OI layers into Si MOSFET channel region is beneficial for enhancing carrier mobilities due to larger separation between carrier sub-bands and reduced electron effective mass.

(2) For bulk-Si N-MOSFETs, gate leakage current is reduced with the OI layers as a result of larger electron tunneling effective mass.

The performance enhancements mentioned above with the OI layers are verified by experiments, as shown in Figure 2.8.

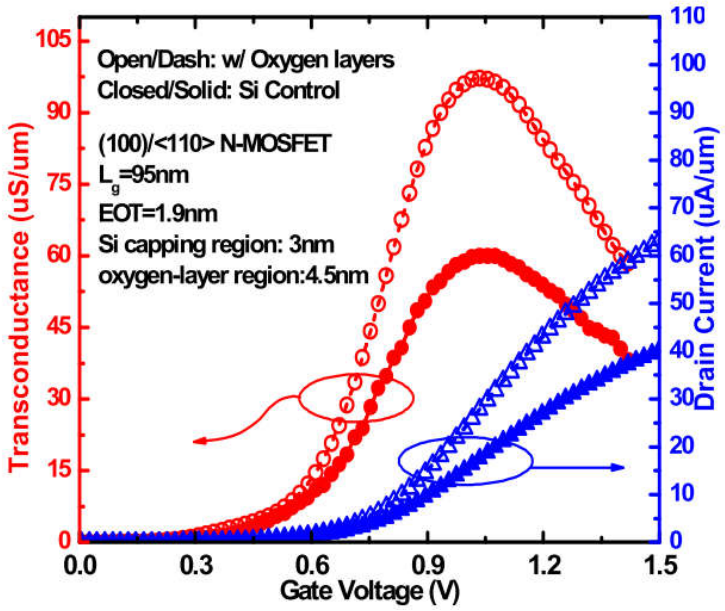

Fig. 2.8. Measured linear current (V_{DS} = 50 mV) and transconductance characteristics for deep submicron (L_g = 95 nm) bulk-Si n-channel MOSFETs with *vs.* without the OI layers.

2.3 Dopant Profile Engineering

Compared with uniform channel doping profiles, super-steep retrograde (SSR) channel doping profiles are beneficial for maximizing current and transconductance values while maintaining device electrostatic integrity to suppress short-channel effects [6, 13, 43, 44]. Typically, boron and phosphorus are implanted into Si to form channel doping profiles for nFETs and pFETs, respectively. However, during the subsequent thermal anneal processes used to activate the dopants, interstitial-induced transient-enhanced diffusion (TED) can strongly enhance boron and phosphorus diffusivity. After the well formation process, gate oxidation and the other implantation (S/D extension implant, and halo implantation) processes also contribute to channel dopant depth profile broadening by injecting interstitials into the well region. To fully achieve the benefits of SSR channel

profiles, it is important to reduce dopant diffusion caused by point defects. One approach is to form Si:C + Si epitaxially-grown channel, which reduces boron diffusion significantly due to the interaction between carbon and boron, achieving SSR channel profiles in nFETs [45, 46]. However, worse device performance ($|V_{TH}|$ increase and $I_{d,sat}$ degradation) are observed for pFETs with the presence of a Si:C epitaxial layer [45]. Considering the importance to achieve good device performance for both nFETs and pFETs, boron and phosphorus TED effects need to be suppressed simultaneously. This is possible with OI technology, which can reduce B and P dopant diffusion while creating a dopant pile-up effect via trapping of interstitials [24, 26, 27, 31].

To study the impacts of the OI layer on interstitials, experiments were performed using p-type (001) crystalline silicon substrates and without the OI layers. First, silicon wafer samples received high-energy boron (B11, 120keV, 1.4×10^{14} cm^{-2}) and phosphorus (P311, 220keV, 8.3×10^{14} cm^{-2}) ion implantations to form p-wells and n-wells, respectively. The samples were subsequently annealed in N_2 ambient at 750°C for 1 hr to activate the implanted dopants. With the presence of the OI layers, boron and phosphorus diffusion towards the silicon substrate surface are effectively suppressed. It can be seen from the SIMS analyses (Figures 2.9, 2.10) that the surface doping concentration values are much lower for the OI samples as compared with the control samples. Because both boron and phosphorus diffusion are interstitial-mediated, this indicates that the OI layers trap or block the diffusion of silicon interstitials. The OI sample comprises multiple partial oxygen monolayers at a depth of around 40 nm. Figures 2.9, 2.10 show boron and phosphorus "pile-up" effects around the depth of the OI region as a result of interstitial trapping. This is not observed for the control samples.

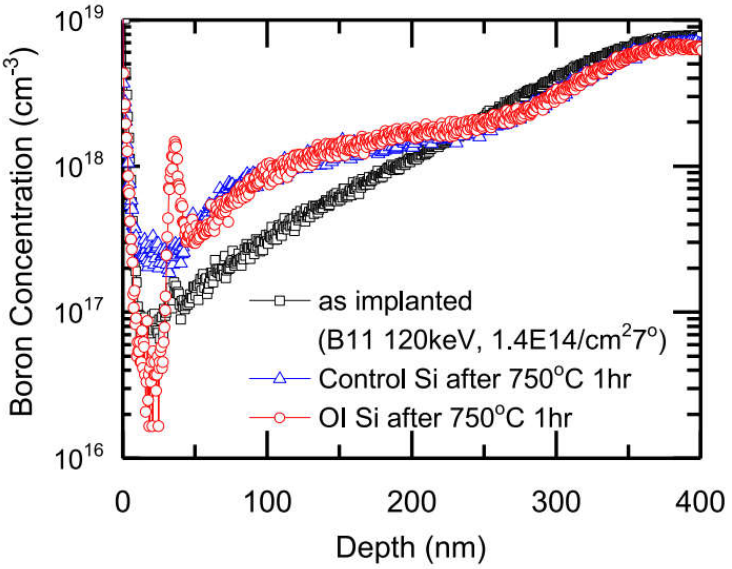

Fig. 2.9. Boron depth profile: as-implanted sample and after anneal (1 hr 750°C N_2 anneal), for control and OI samples.

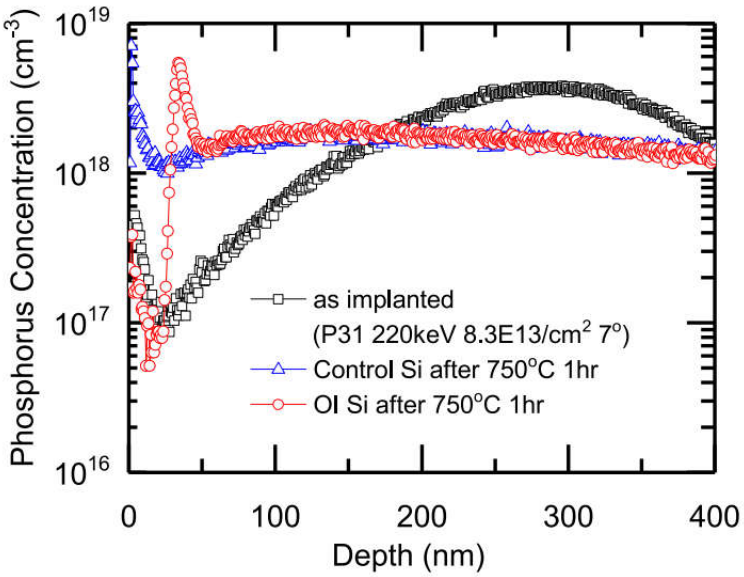

Fig. 2.10. Phosphorus depth profile: as-implanted sample and after anneal (1 hr 750°C N_2 anneal), for control and OI samples.

To investigate the impacts of the OI layers on interstitials injection caused by silicon oxidation, control and OI samples were prepared by implanting B ions with an acceleration energy of 25 keV, at a dose of 2×10^{13} cm^{-2}, through an 80 Å thick screening oxide layer. The OI sample comprises multiple partial oxygen monolayers inserted into the crystalline Si substrate at a depth of around 40 nm. A 5s 1050 °C rapid thermal anneal (RTA) was carried out in a conventional lamp-heated RTA tool (AccuThermo model AW610) to activate the implanted B atoms. Then the samples were cleaned to remove the screening oxide and received an oxidizing anneal for 60 minutes at 800 °C to grow 125 Å thick oxide, followed by an 850 °C, 30 minutes post-oxidation anneal. Finally, the samples were subjected to a 1000 °C, 2 minutes RTA to mimic the source/drain activation anneal process. SIMS measurements (Figures 2.11, 2.12) show that, with the OI layers, the surface boron concentration value stays within the $1 - 2 \times 10^{16}$ cm^{-3} range after the thermal oxidation and the RTA treatment. Whereas for the control sample, the surface boron concentration reaches mid-10^{17} cm^{-3} after thermal treatments. More strikingly, the as-implanted boron doping profile shape beyond the depth of 40 nm is retained with the OI layers due to trapping of interstitials injected into the silicon substrate during the surface oxidation process.

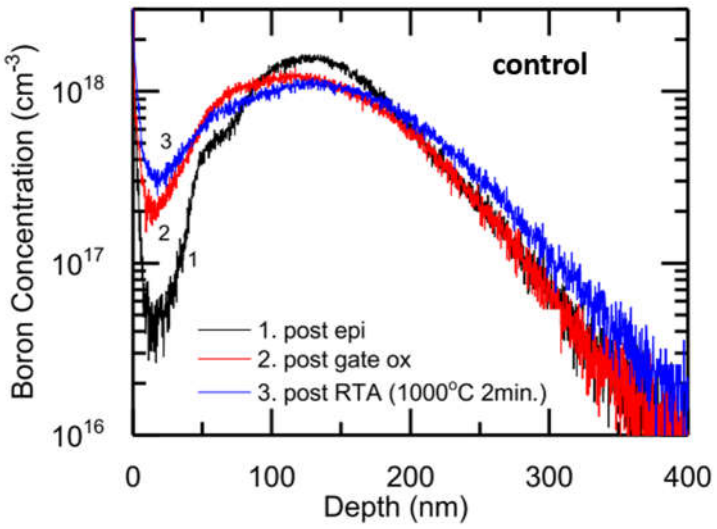

Fig. 2.11. Boron depth profiles for: 1. the as-implanted control sample. 2. the control sample after receiving an oxidizing 800 °C 60 minutes anneal and an 850 °C 30 minutes N_2 anneal. 3. the control sample after receiving a 1000 °C 2 minutes RTA [24].

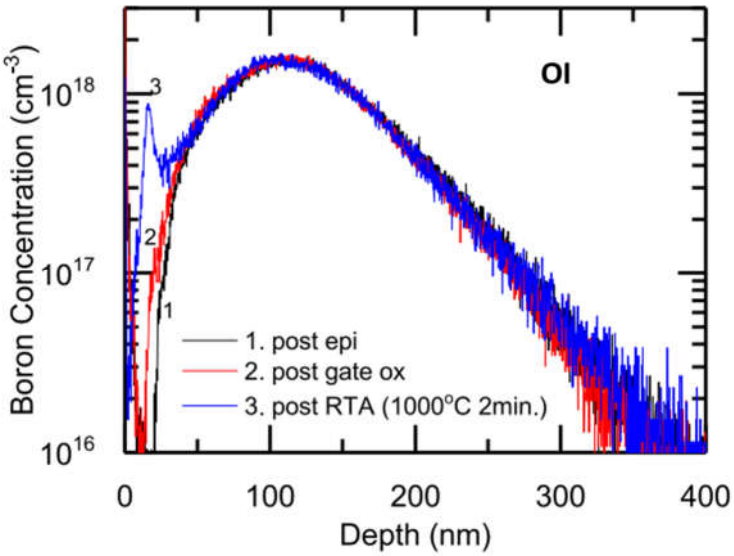

Fig. 2.12. Boron depth profiles for: 1. the as-implanted OI sample. 2. the OI sample after receiving an oxidizing 800 °C 60 minutes anneal and an 850 °C 30 minutes N_2 anneal. 3. the OI sample after receiving a 1000 °C 2 minutes RTA [24].

To verify that the enhanced boron diffusion shown in Figure 2.11 cannot be attributed to implant-induced damage, in-situ formed boron doped layers were used to verify that the OI layers can reduce boron oxidation-enhanced diffusion (OED) [31]. Figure 2.13 explains schematically how two 10-nm thick boron marker layers each with a boron dose

of 2.6 × 10^{14} cm^{-2} are inserted between 45-nm thick undoped buffer Si layers. In the OI sample, a 6 nm thick OI layer is incorporated in-between the two B marker layers. Then the samples were subjected to an 800 ^{o}C dry oxidation anneal for 30 minutes. Since B atoms are incorporated in-situ, it can be expected that no point defects were introduced during the epitaxy process to grow B marker layers. Thus, B profile broadening can be reasonably attributed to the interstitials introduced during the thermal oxidation step. SIMS measurements were used to compare B diffusions in control and OI samples. It can be seen from a comparison of Figures 2.14, 2.15 that with the OI layers, the buried (lower) B marker layer experienced little to no diffusion. The interstitials injected by surface oxidation are believed to be trapped around the OI layers, which also causes the observed B "pile-up" effect shown in Figure 2.15.

Fig. 2.13. Cross-section schematics of boron marker layers in epitaxially grown silicon samples. 6 nm thick OI layers protect the lower boron marker layer from OED.

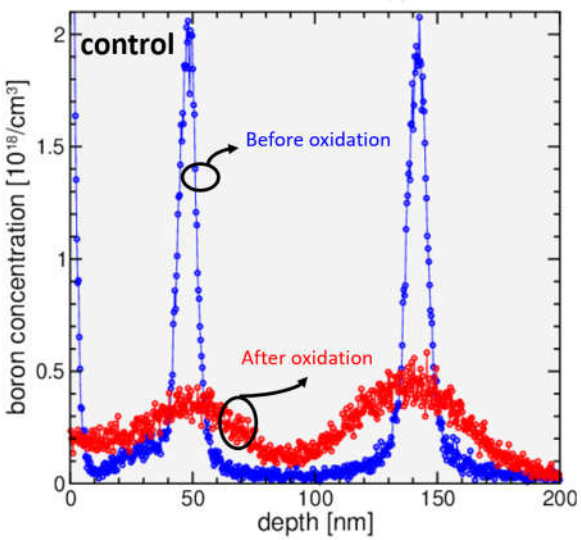

Fig. 2.14. Boron depth profiles for the control sample. Both B marker layers experienced substantial diffusion due to oxide-enhanced-diffusion [31].

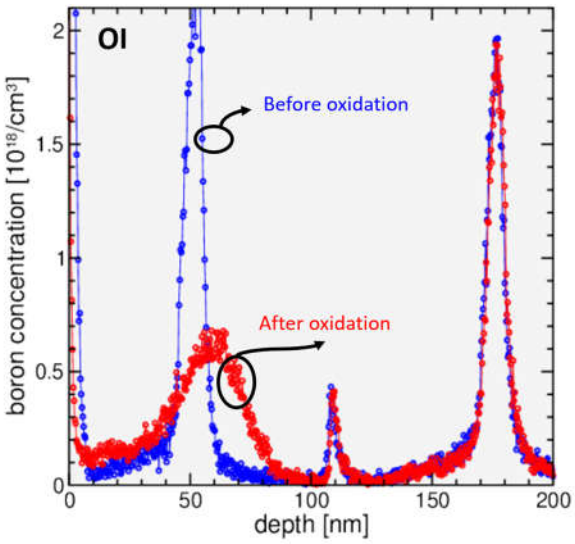

Fig. 2.15. Boron depth profiles for the OI sample. The buried B peak retained its original shape because of suppressed OED due to the presence of the OI layers [31].

Buried OI layers are experimentally verified to reduce TED and OED. This indicates that OI technology is a promising candidate to facilitate SSR channel profile formation, which can enhance device performance while maintaining superior electrostatic integrity.

2.4 Summary

In this chapter, OI technology is introduced and its benefits for improving MOSFET performance are discussed. OI layers are verified experimentally to improve e- and h+ mobility and reduce gate leakage due to carrier sub-band engineering. In addition, they are beneficial for reducing TED and OED by trapping of Si interstitials and hence are projected to help form SSR channel dopant profile, which also is beneficial for boosting carrier mobility and suppress SCE.

2.5 References

[1] G. Baccarani, M. R. Wordeman, and R. H. Dennard, "Generalized scaling theory and its application to a ¼ micrometer MOSFET design," *IEEE T Electron Dev,* vol. 31, no. 4, pp. 452-462, 1984.

[2] M. Bohr, "A 30 year retrospective on Dennard's MOSFET scaling paper," *IEEE Solid-State Circuits Society Newsletter,* vol. 12, no. 1, pp. 11-13, 2007.

[3] K. K. Young, "Short-channel effect in fully depleted SOI MOSFETs," *IEEE T Electron Dev,* vol. 36, no. 2, pp. 399-402, 1989.

[4] T. Ghani *et al.*, "Scaling challenges and device design requirements for high performance sub-50 nm gate length planar CMOS transistors," in *Symposium on VLSI*

Technology, Honolulu, HI, USA, 13-15 June 2000 2000: IEEE, doi: 10.1109/VLSIT.2000.852814.

[5] A. Chaudhry and M. J. Kumar, "Controlling short-channel effects in deep-submicron SOI MOSFETs for improved reliability: a review," *IEEE Transactions on Device and Materials Reliability,* vol. 4, no. 1, pp. 99-109, 2004.

[6] D. A. Antoniadis and J. E. Chung, "Physics and technology of ultra short channel MOSFET devices," in *International Electron Devices Meeting 1991 [Technical Digest]*, 1991: IEEE, pp. 21-24.

[7] X. Zhang *et al.*, "Physical model of the impact of metal grain work function variability on emerging dual metal gate MOSFETs and its implication for SRAM reliability," in *2009 IEEE International Electron Devices Meeting (IEDM)*, 2009: IEEE, pp. 1-4.

[8] X. Wang, B. Cheng, A. R. Brown, C. Millar, and A. Asenov, "Statistical variability in 14-nm node SOI FinFETs and its impact on corresponding 6T-SRAM cell design," in *2012 Proceedings of the European Solid-State Device Research Conference (ESSDERC)*, 2012: IEEE, pp. 113-116.

[9] X. Wang, A. R. Brown, B. Cheng, and A. Asenov, "Statistical variability and reliability in nanoscale FinFETs," in *2011 International Electron Devices Meeting*, 2011: IEEE, pp. 5.4. 1-5.4. 4.

[10] T. Matsukawa *et al.*, "Comprehensive analysis of variability sources of FinFET characteristics," in *2009 Symposium on VLSI Technology*, 2009: IEEE, pp. 118-119.

[11] K. El Sayed, E. Lyumkis, and A. Wettstein, "Modeling statistical variability with the impedance field method," in *Proc. Int. Conf. Simulation Semiconductor Process. Devices (SISPAD)*, 2012, pp. 205-208.

[12] A. Asenov, "Simulation of statistical variability in nano MOSFETs," in *2007 IEEE symposium on VLSI technology*, 2007: IEEE, pp. 86-87.

[13] I. De and C. M. Osburn, "Impact of super-steep-retrograde channel doping profiles on the performance of scaled devices," *IEEE T Electron Dev,* vol. 46, no. 8, pp. 1711-1717, 1999.

[14] D. C. Boyd, J. R. Holt, M. Ieong, R. T. Mo, Z. Ren, and G. G. Shahidi, "Ultra-thin body super-steep retrograde well (SSRW) FET devices," USA Patent Appl. 10/710,736, 2006.

[15] H. Irie, K. Kita, K. Kyuno, and A. Toriumi, "In-plane mobility anisotropy and universality under uni-axial strains in nand p-MOS inversion layers on (100),[110], and (111) Si," in *IEDM Technical Digest. IEEE International Electron Devices Meeting, 2004.*, 2004: IEEE, pp. 225-228.

[16] M. Chu, Y. Sun, U. Aghoram, and S. E. Thompson, "Strain: A solution for higher carrier mobility in nanoscale MOSFETs," *Annual Review of Materials Research,* vol. 39, pp. 203-229, 2009.

[17] C. Wu *et al.*, "High performance 22/20nm FinFET CMOS devices with advanced high-K/metal gate scheme," in *2010 International Electron Devices Meeting*, 2010: IEEE, pp. 27.1. 1-27.1. 4.

[18] C. Auth *et al.*, "A 22nm high performance and low-power CMOS technology featuring fully-depleted tri-gate transistors, self-aligned contacts and high density MIM capacitors," in *2012 Symposium on VLSI Technology (VLSIT)*, 2012: IEEE, pp. 131-132.

[19] N. Xu *et al.*, "Electron mobility enhancement in (100) oxygen-inserted silicon channel," *Applied Physics Letters,* vol. 107, no. 12, Sep 21 2015.

[20] N. Xu *et al.*, "Oxygen-inserted SegFET: A candidate for 10-nm node system-on-chip applications," in *2014 Silicon Nanoelectronics Workshop (SNW)*, 2014: IEEE, pp. 1-2.
[21] N. Xu *et al.*, "Extension of Planar Bulk n-Channel MOSFET Scaling With Oxygen Insertion Technology," *IEEE T Electron Dev,* vol. 61, no. 9, pp. 3345-3349, 2014, doi: 10.1109/TED.2014.2342496.
[22] N. Xu *et al.*, "Effectiveness of Quasi-Confinement Technology for Improving P-Channel Si and Ge MOSFET Performance."
[23] N. Xu *et al.*, "MOSFET performance and scalability enhancement by insertion of oxygen layers," in *2012 International Electron Devices Meeting*, 2012: IEEE, pp. 6.4. 1-6.4. 4.
[24] H. Takeuchi *et al.*, "Punch-Through Stop Doping Profile Control via Interstitial Trapping by Oxygen-Insertion Silicon Channel - IEEE Journals & Magazine," *IEEE Journal of the Electron Devices Society,* vol. 6, pp. 481-486, 2018, doi: 10.1109/JEDS.2017.2769682.
[25] J. Smith *et al.*, "Experimental Investigation of N-Channel Oxygen-Inserted (OI) Silicon Channel MOSFETs with High-K/Metal Gate Stack," in *2018 76th Device Research Conference (DRC)*, 2018: IEEE, pp. 1-2.
[26] R. J. Mears *et al.*, "Silicon superlattice on SOI for high mobility and reduced leakage," in *2007 IEEE International SOI Conference*, 2007: IEEE, pp. 23-24.
[27] R. Mears *et al.*, "Simultaneous carrier transport enhancement and variability reduction in Si MOSFETs by insertion of partial monolayers of oxygen," in *2012 IEEE Silicon Nanoelectronics Workshop (SNW)*, 2012: IEEE, pp. 1-2.
[28] A. Marshall *et al.*, "The impact of oxygen insertion technology on SRAM yield performance," in *2017 IEEE Electron Devices Technology and Manufacturing Conference (EDTM)*, 2017: IEEE, pp. 191-192.
[29] S. Halilov *et al.*, "Tunneling through superlattices: the effect of anisotropy and kinematic coupling," *Journal of Physics: Condensed Matter,* vol. 24, no. 49, p. 495801, 2012.
[30] N. Damrongplasit *et al.*, "Comparative study of uniform versus supersteep retrograde MOSFET channel doping and implications for 6-T SRAM yield," *IEEE T Electron Dev,* vol. 60, no. 5, pp. 1790-1793, 2013.
[31] D. Connelly *et al.*, "Suppressing oxidation-enhanced diffusion of boron via buried epitaxial oxygen-inserted layers in silicon," in *2018 IEEE 2nd Electron Devices Technology and Manufacturing Conference (EDTM)*, 2018: IEEE, pp. 163-165.
[32] K.-Y. Lim *et al.*, "Novel stress-memorization-technology (SMT) for high electron mobility enhancement of gate last high-k/metal gate devices," in *2010 International Electron Devices Meeting*, 2010: IEEE, pp. 10.1. 1-10.1. 4.
[33] N. Xu, B. Ho, M. Choi, V. Moroz, and T.-J. K. Liu, "Effectiveness of stressors in aggressively scaled FinFETs," *IEEE T Electron Dev,* vol. 59, no. 6, pp. 1592-1598, 2012.
[34] S. Eneman *et al.*, "Layout impact on the performance of a locally strained PMOSFET," in *Digest of Technical Papers. 2005 Symposium on VLSI Technology, 2005.*, 2005: IEEE, pp. 22-23.
[35] N. Xu, L. T.-N. Wang, A. R. Neureuther, and T.-J. K. Liu, "Physically based modeling of stress-induced variation in nanoscale transistor performance," *IEEE Transactions on Device and Materials Reliability,* vol. 11, no. 3, pp. 378-386, 2011.

[36] S. J. Clark *et al.*, "First principles methods using CASTEP," *Zeitschrift für Kristallographie-Crystalline Materials,* vol. 220, no. 5/6, pp. 567-570, 2005.
[37] M. Segall *et al.*, "First-principles simulation: ideas, illustrations and the CASTEP code," *Journal of Physics: Condensed Matter,* vol. 14, no. 11, p. 2717, 2002.
[38] E. Ungersboeck, S. Dhar, G. Karlowatz, V. Sverdlov, H. Kosina, and S. Selberherr, "The effect of general strain on the band structure and electron mobility of silicon," *IEEE T Electron Dev,* vol. 54, no. 9, pp. 2183-2190, 2007.
[39] M. Fischetti, Z. Ren, P. Solomon, M. Yang, and K. Rim, "Six-band k· p calculation of the hole mobility in silicon inversion layers: Dependence on surface orientation, strain, and silicon thickness," *Journal of Applied Physics,* vol. 94, no. 2, pp. 1079-1095, 2003.
[40] D. Esseni and A. Abramo, "Modeling of electron mobility degradation by remote Coulomb scattering in ultrathin oxide MOSFETs," *IEEE T Electron Dev,* vol. 50, no. 7, pp. 1665-1674, 2003.
[41] S.-i. Takagi, A. Toriumi, M. Iwase, and H. Tango, "On the universality of inversion layer mobility in Si MOSFET's: Part I-effects of substrate impurity concentration," *IEEE T Electron Dev,* vol. 41, no. 12, pp. 2357-2362, 1994.
[42] E. B. Ramayya, D. Vasileska, S. M. Goodnick, and I. Knezevic, "Electron mobility in silicon nanowires," *IEEE Transactions on Nanotechnology,* vol. 6, no. 1, pp. 113-117, 2007.
[43] S. Thompson, P. Packan, and M. Bohr, "Linear versus saturated drive current: Tradeoffs in super steep retrograde well engineering," in *1996 Symposium on VLSI Technology. Digest of Technical Papers*, 1996: IEEE, pp. 154-155.
[44] A. Asenov and S. Saini, "Suppression of random dopant-induced threshold voltage fluctuations in sub-0.1μm MOSFET's with epitaxial and δ-doped channels," *IEEE T Electron Dev,* vol. 46, no. 8, pp. 1718-1724, 1999.
[45] A. Hokazono, H. Itokawa, I. Mizushima, S. Kawanaka, S. Inaba, and Y. Toyoshima, "Steep channel profiles in n/pMOS controlled by boron-doped Si:C layers for continual bulk-CMOS scaling," in *2009 IEEE International Electron Devices Meeting (IEDM)*, 2009: IEEE, pp. 1-4.
[46] A. Hokazono *et al.*, "Steep channel & Halo profiles utilizing boron-diffusion-barrier layers (Si:C) for 32 nm node and beyond," in *2008 Symposium on VLSI Technology*, 2008: IEEE, pp. 112-113.

Chapter 3

Oxygen Insertion Technology for Ultra-shallow Junction Formation

In this chapter, the effects of oxygen insertion (OI) technology, a low-temperature deposited oxide (LTO) capping layer, and a SiN_x capping layer on ultra-shallow junction (USJ) formation are discussed [1]. OI technology is demonstrated to be beneficial for reducing junction depth (X_J) and mitigating sheet resistance (R_{sh}) increase due to retarded interstitial-driven diffusion and enhanced dose retention during thermal annealing [2].

3.1 Introduction

Low resistivity ultra-shallow source/drain (S/D) extension regions are necessary for short-channel planar CMOSFETs (Complementary Metal Oxide Semiconductor Field Effect Transistors) [3, 4]. The depth of the source/drain extension junctions (X_J) must be smaller than transistor gate length (L_{gate}) to suppress short-channel effects effectively [3, 5]. Also, a high level of dopant activation in the S/D extension regions is required to achieve low S/D parasitic series resistances [4-6]. Typically, ultra-low energy implants followed by a rapid thermal anneal (RTA) treatment is the preferred method to form ultra-shallow junctions. Due to the point defects created by the implantation, transient-enhanced diffusion (TED) during the RTA treatment results in a diffusion "tail" for boron (B) [7, 8] and "kink-and-tail" for phosphorus (P) [9, 10]. Therefore, as L_{gate} scales down, source/drain junction depths do not scale commensurately. Another major challenge associated with the ultra-low energy implantation technique is dopant clustering [11, 12] or precipitation [13, 14], limiting the electrical activation level of dopants that can be achieved. Consequently, as X_J scales, S/D parasitic resistances increase significantly, limiting the short-channel MOSFET performance improvement brought by reducing X_J. These challenges need to be overcome to maximize the performance of advanced planar CMOS transistors.

A previous study of super-steep retrograde (SSR) channel dopant profile formation shows the efficacy of oxygen partial monolayers [15, 16] for reducing B and P diffusion, improving field effect mobility and suppressing random-dopant fluctuations [17]. In this work, the impacts of the OI layers on dopant diffusion and activation for USJ formation are investigated by experiments and technology computer-aided design (TCAD) modeling.

Additionally, the efficacy of an oxide capping layer and a SiN_x capping layer on dopant diffusion is studied for the first time. The effects of an oxide capping layer on dopant electrical activation in USJs are also discussed.

3.2 Ultra-shallow Junction Formation

Ultra-shallow p+/n and n+/p junctions were formed in (100) p-type control and OI substrates [1]. In OI substrates, multiple partial monolayers of oxygen were incorporated into crystalline silicon at a depth of approximately 10nm beneath the surface.

Dopant atoms were implanted through a thin screening oxide layer (2 nm SiO_2) into p-type control and OI substrates with 0-degree tilt/rotation. The oxide screening layer also serves to protect the surface from contamination during ion implantation. For the OI layers to effectively block dopant diffusion into the substrate, the implantation conditions are chosen according to the stopping and range of ions in matter (SRIM) simulations [18] such that the projected range is less than 10 nm. Table 3.1 summarizes ion implantation conditions. Because only p-type substrates are available, B samples were first counter-doped with a 90keV, 2×10^{13}cm^{-2} arsenic implant, imitating an n-type well formation process [19]. The As well formation made R_{sh} extraction possible for B samples. Figure 3.1 shows depth profiles of dopant atoms and oxygen atom obtained by secondary ion mass spectroscopy (SIMS) after ion implantation.

Sample	Implant Species	Implant Energy (keV)	Dose (cm^{-2})
Boron	BF_2	2.5	1×10^{15}
Phosphorus	P	1	1×10^{15}
Arsenic	As	1	1×10^{15}

Table 3.1. Ion Implantation conditions used in this work.

After the ion implantation, the wafers were broken into 2×2 cm^2 pieces. The samples were first cleaned in a sulfuric peroxide mixture (H_2SO_4: H_2O_2) bath at 120 ^{o}C for 10 minutes followed by a 1-min dip in the dilute hydrofluoric acid (DHF) solution (10:1 H_2O:49% HF) until the surface became hydrophobic. The samples were then immediately rinsed in de-ionized (DI) water before receiving a spike anneal in inert N_2 ambient at 1050 ^{o}C using a conventional lamp-heated RTA tool (AccuThermo model AW610). The temperature was first stabilized at 700 ^{o}C for 10 s and subsequently ramped to 1050 ^{o}C

for a duration less than 1 s. It is worth noting that previous work has shown a fast ramp-up rate (300 ^{o}C/s) is beneficial for suppressing TED-induced effects [7]. However, such high ramp-up rate is not possible with a conventional RTA tool to avoid temperature over-shooting. In this work, the ramp-up rate was 116 ^{o}C/s to ensure temperature stability.

In OI substrates, each oxygen atom is inserted interstitially into the crystalline silicon substrate and bonds with two neighboring silicon atoms [15]. From the SIMS measurements of oxygen concentration profile shown in Figure 3.2, it can be seen that the Si-O bond configuration prevents the OI layers from decomposing during the 1050^{o}C spike anneal. The excellent thermal stability of the partial oxygen monolayers allows OI technology to be readily integrated into the standard CMOS process flow.

In addition to shallow X_J, low resistivity (*i.e.,* high conductivity) is another prominent figure of merit for S/D extension engineering [4]. Previous findings show a significant amount of dose loss for ultra-shallow junctions during low-temperature (700 ^{o}C ~ 900 ^{o}C) anneals due to out-diffusion effects [20-22]. It was hypothesized that the free Si surface serves to act as either a sink for interstitials or a trap for dopant atoms. Other researchers found significant R_{sh} increase after stripping native oxide off the substrate after annealing [23]. This was attributed to dopant segregation effects associated with the Si/SiO_2 interface. To examine the correlation between the interface and the "dose loss" effect, some samples were capped with an insulating layer prior to RTA treatment, either composing of 10nm-thick low-temperature oxide (LTO) formed at 400^{o}C or 10nm-thick silicon nitride (SiN_x) formed at 200^{o}C by chemical vapor deposition (CVD). After the 1050 ^{o}C spike anneal, the capping layer was stripped in DHF before SIMS analysis.

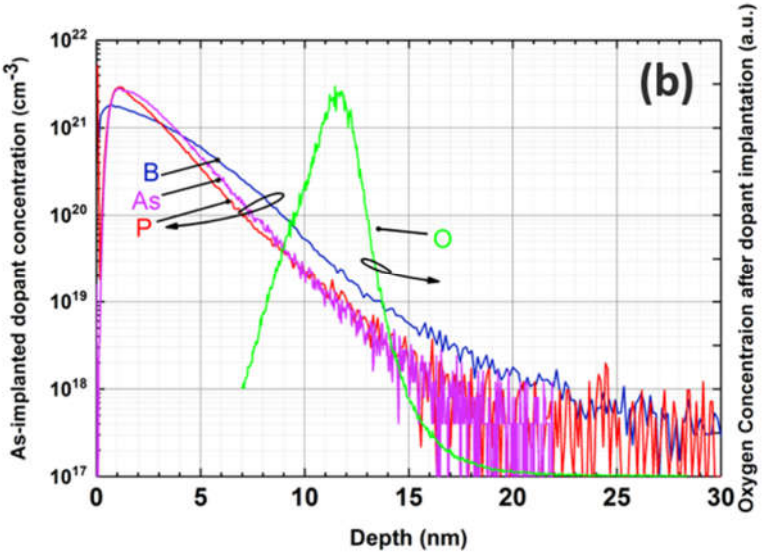

Fig. 3.1. Depth concentration profiles of dopant (log scale) and oxygen atoms (linear scale) from SIMS measurements after ion implantation. Notice that for B, As, and P-implanted samples, the projected range is shallower than the region containing oxygen monolayers.

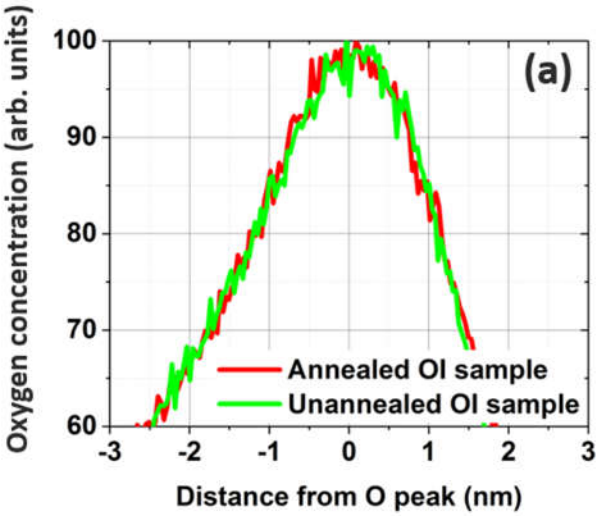

Fig. 3.2. Depth concentration profiles of oxygen before and after the 1050 °C spike anneal from SIMS measurements. The comparison shows that oxygen monolayers can withstand high-temperature RTA treatment.

3.3 Dose Loss and Diffusion Analysis

SIMS was used to study the effects of the capping layer and the OI layers on dopant dose loss and diffusion. Total dopant dose is calculated by integrating the dopant concentration profile from SIMS measurements.

3.3.1 B Dose Loss and Diffusion Analysis

Table 3.2 provides a summary of total B dose in the various B-doped samples. The comparison between uncapped and capped samples indicates that neither LTO nor SiN_x capping layer helps retain B dose during the 1050°C spike anneal. It was found that 39% of the implanted B was lost from the annealed uncapped control sample, whereas only 16% of the implanted B was lost from the annealed uncapped OI sample. A higher B dose retained suggests the OI layers shall be beneficial for achieving low-resistivity p-type ultra-shallow junctions.

B Sample	B dose ($\times 10^{14}$ cm^{-2})	Dose loss (%)
As-implanted	7.4	0
Uncapped control	4.5	39
LTO capped control	3.6	51
SiN_x capped control	3.7	50
Uncapped OI	6.2	16
LTO capped OI	4.7	36
SiN_x capped OI	4.6	38

Table 3.2. Boron dose in control and OI samples.

Figure 3.2 plots the B depth profiles in annealed control samples. C_{enh} is defined as the concentration below which B diffusion is enhanced. The SIMS data shows that C_{enh} is approximately 1×10^{20} cm^{-3}, consistent with previous findings [7]. C_{enh} is dependent on the intrinsic carrier concentration during the annealing process and thus can be expected to be not affected by the presence of a capping layer. Similar surface B concentration values were found in all three annealed samples, suggesting negligible B segregation to the interface between the capping layer and the silicon substrate. The enhanced B diffusion in capped samples can be explained by strain-induced interstitial generation due to the capping layer since B diffusion is mainly interstitial-driven [24, 25].

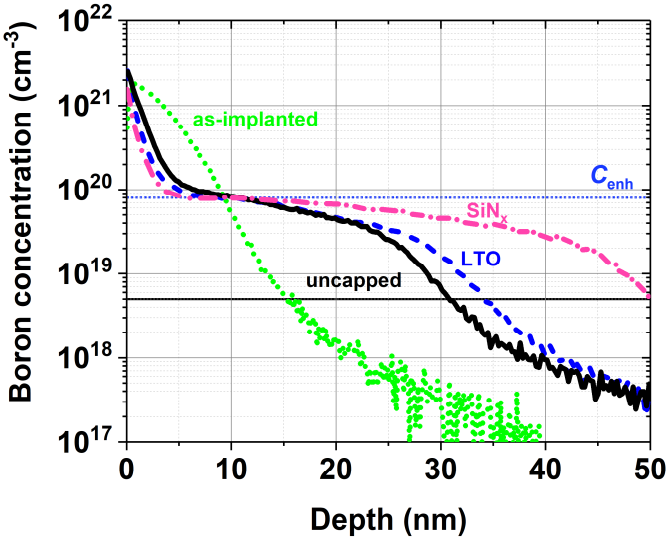

Fig. 3.2. Depth concentration profiles of Boron in control Si from SIMS measurements. C_{enh} is the same for the uncapped and the capped samples. Notice that B diffusion was enhanced in both LTO or SiN_x capped sample.

Significant B uphill diffusion was found in all three annealed control samples. Previous work on ultra-low energy B implants into bare silicon substrates revealed similar uphill diffusion phenomenon during the low temperature (700°C ~ 900°C) RTA treatments [20, 26]. It was hypothesized that the bare silicon substrate surface acts as a sink for interstitials, driving B out-diffusion since B diffusion is primarily interstitial-driven. Since ubiquitous B uphill diffusion was observed in both uncapped and capped samples, this indicates the interface between the capping layer and the substrate can also act as a sink for interstitials.

Since the peak retrograde well dopant concentration is in the range of 10^{18} cm^{-3} in advanced planar CMOS devices, the nominal junction depth is taken to be the depth at which the dopant concentration falls to 5×10^{18} cm^{-3}. To compare X_J fairly, Figure 3.3 shows the B depth profiles for a control sample and the OI samples with a similar retained dopant dose as compared with that for the control sample. From the SIMS measurements, it is deduced that B diffusion is effectively reduced beyond the OI layers, possibly due to the impeded diffusion of silicon self-interstitials beyond the OI layers since O atoms are inserted into crystalline silicon interstitially. Prior studies of B diffusion [27] showed that C_{enh} is dependent on the annealing temperature. Nevertheless, it can be seen that C_{enh} values are higher in OI samples. The higher plateau at a depth of 10 nm is most likely caused by the OI layers, which are believed to cause a dopant pile-up effect, in addition to reducing B diffusion.

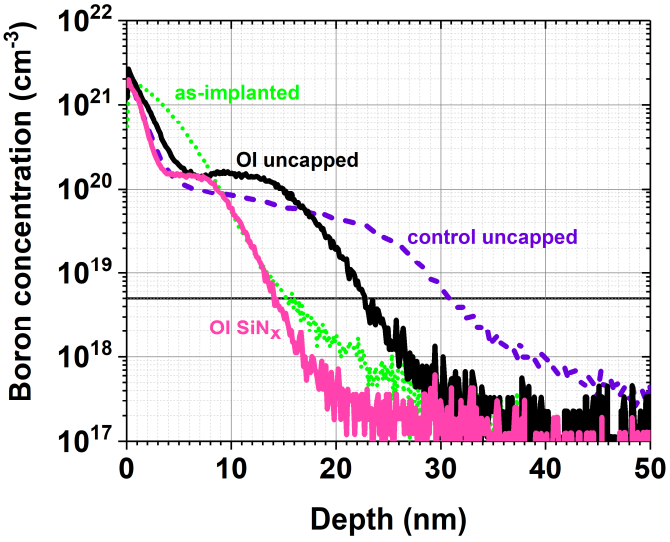

Fig. 3.3. Comparison of B depth profiles for control *vs.* OI samples from SIMS measurements. For the uncapped control sample, the retained B dose is 4.5×10^{14} cm^{-2} and $X_J = \mathbf{31\ nm}$. For the SiN$_x$ capped **OI sample**, the retained B dose is 4.6×10^{14} cm^{-2} and $X_J = \mathbf{14\ nm}$. For the uncapped **OI sample**, the retained B dose is even higher, 6.2×10^{14} cm^{-2}, but the junction depth ($X_J = \mathbf{22\ nm}$) is still shallower than for the uncapped control sample.

3.3.2 P Dose Loss and Diffusion Analysis

Table 3.3 summarizes the total P dose in the various P-doped samples. At least 70% of the implanted P was lost from the annealed uncapped samples, as compared with the as-implanted sample. The comparison between uncapped and capped samples suggests that both LTO and SiN_x capping layer helped to retain P dose during the RTA treatment.

Figure 3.4 plots the P depth profiles in control samples. Previous studies of P USJ formation found P uphill diffusion and dose loss as a result of the bare Si surface acting as a sink for interstitials. The steep P profile gradient near the surface indicates P uphill diffusion in the uncapped sample. However, it can be seen that there is less P out-diffusion in the capped samples. This is likely caused by substantially enhanced diffusion into the substrate due to the aforementioned strain-induced interstitial generation associated with the capping layer.

Figure 3.5 provides a comparison of P depth profiles for a control sample and an OI sample with a similar retained P dose. P diffusion beyond the OI layers is reduced effectively, reducing X_J by 47% in the OI sample as compared with that for the control sample. This is likely due to the aforementioned impeded diffusion of silicon self-interstitials. Similar to as for B, a dopant pile-up effect in the region of the OI layers is observed for P.

P Sample	P dose ($\times 10^{14}$ cm^{-2})	Dose loss (%)
As-implanted	7.5	0
Uncapped control	2.2	70
LTO capped control	3.6	52
SiN_x capped control	3.2	57
Uncapped OI	1.6	78
LTO capped OI	3.4	55
SiN_x capped OI	4.1	45

Table 3.3. P dose in control and OI samples.

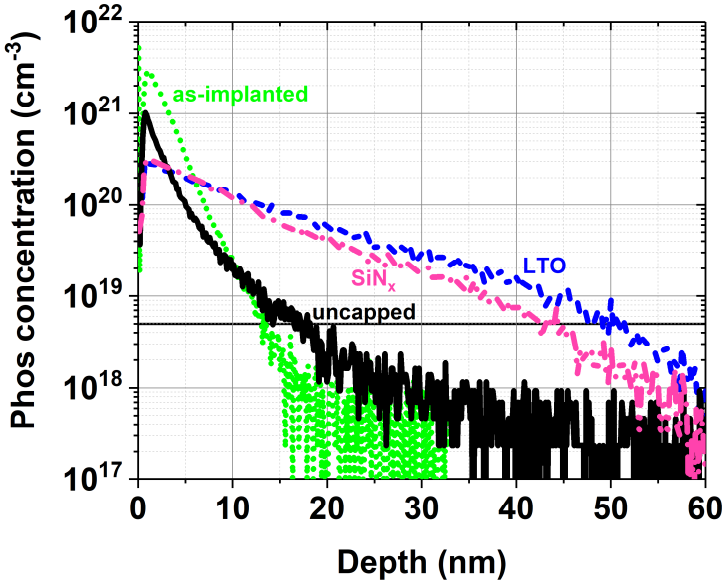

Fig. 3.4. Depth concentration profiles of Phosphorus in control Si from SIMS measurements. C_{enh} is the same for the uncapped and the capped samples. Notice that uphill diffusion was observed only in the uncapped sample. Notice that in the case of LTO or SiN_x capped samples, P diffusion was enhanced significantly.

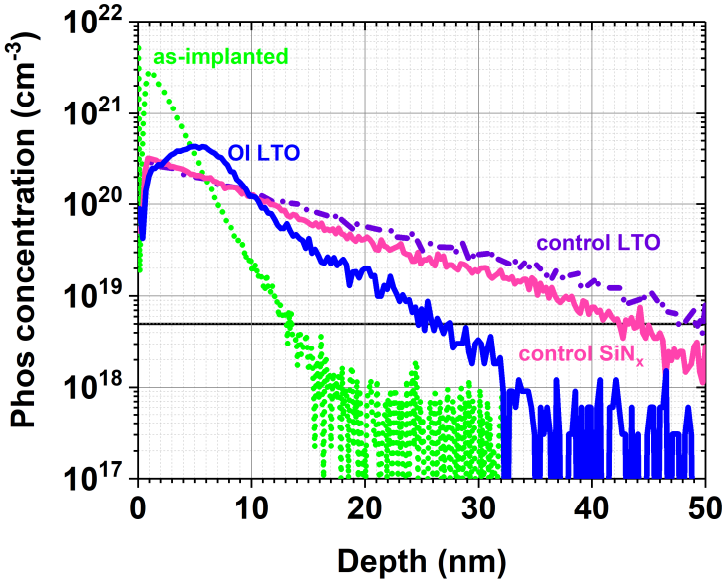

Figure 3.5. Comparison of P depth profiles for control *vs.* OI samples from SIMS measurements. For the LTO-capped **control sample**, the retained P dose is 3.6×10^{14} cm^{-2} and $X_J = \mathbf{49\ nm}$. For the LTO-capped **OI sample**, the retained P dose is 3.4×10^{14} cm^{-2} and $X_J = \mathbf{26\ nm}$. Notice that in the case of the OI sample, X_J was reduced by about 50% with a similar P dose retained.

3.3.3 As Dose Loss and Diffusion Analysis

Table 3.4 summarizes the total As dose in the various arsenic-doped samples. It can be seen that more than 70% of the implanted As was lost from the annealed uncapped control sample, as compared with the as-implanted sample. Only SiN_x capping layer helped to retain As dose during the RTA treatment. However, SiN_x capping layer induced dose retention was found to be negligible for OI samples. 51% of the implanted As was lost from both uncapped and SiN_x capped OI samples during the RTA treatment. The comparison between control and OI samples indicates the OI layers helped to retain As during the RTA treatment, suggesting that oxygen insertion technology should be promising for achieving low-resistivity As (n-type) ultra-shallow junction.

Figure 3.6 shows the arsenic depth profiles in the control samples. Thanks to the lower diffusivity of As *vs.* B and P, it can be seen that the As profiles were very shallow (note the smaller depth scale as compared with Figs 3.2, 3.4). The presence of a capping layer does not significantly enhance As diffusion since As diffusion at 1050 ^{o}C is only 40% interstitial-driven.

As Sample	As dose ($\times 10^{14}$ cm^{-2})	Dose loss (%)
As-implanted	8.2	0
Uncapped control	2.4	71
LTO capped control	2.3	72
SiN_x capped control	4.3	48
Uncapped OI	4.0	51
LTO capped OI	3.0	63
SiN_x capped OI	4.0	51

Table 3.4. As dose in control and OI samples.

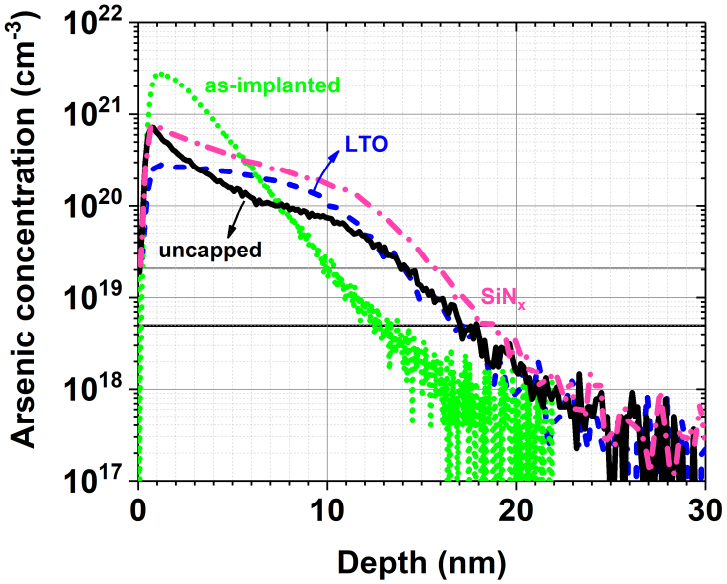

Fig. 3.6. Depth concentration profiles of arsenic in control Si from SIMS measurements.

Figure 3.7 provides a comparison of As depth profiles for a control sample and an OI sample with a similar retained As dose. Since As diffusion is primarily vacancy-driven at 1050 °C and the OI layers reduce B and P diffusion by impeding the diffusion of silicon self-interstitials, it can be expected that the impact of the OI layers on As diffusion is less than that for B and P. It can be seen that X_J is reduced by 16% in the OI sample.

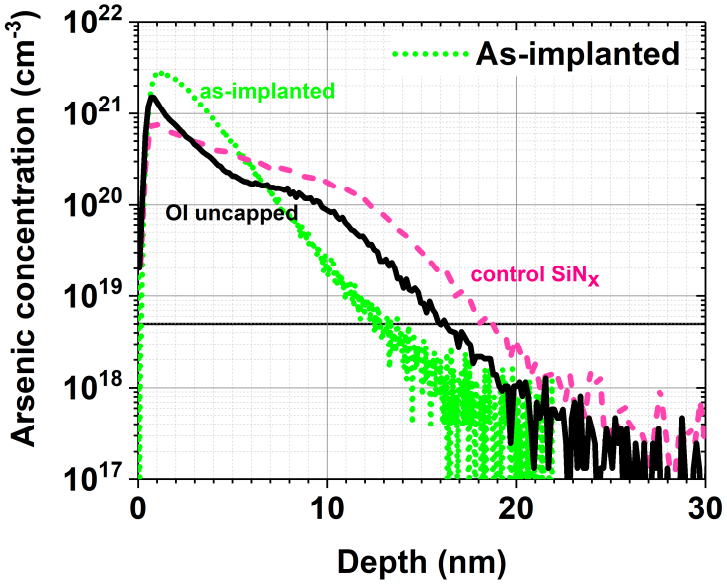

Fig. 3.7. Comparison of As profiles for control *vs.* OI samples. For the SiN_x-capped **control sample,** the retained As dose is 4.3×10^{14} cm^{-2} and X_J = **19 nm**. For the uncapped **OI sample**, the retained As dose is 4.0×10^{14} cm^{-2} and X_J = **16 nm**.

3.4 Sheet Resistance and Dopant Activation Analysis

When investigating the formation of ultra-shallow junctions, it is not enough to just consider X_J. Sheet resistance or the electrical activation of dopants should also be taken into account. A high electrically active dopant concentration in the heavily doped S/D extension regions is desirable for achieving low parasitic S/D resistances. Since SIMS only provides high-resolution dopant depth data, R_{sh} measurements must be performed to determine the impact of oxygen-inserted layers on dopant activation.

Historically, four-point probe [28] and spreading resistance profiling [29] are typically used for electrical characterization of p-n junctions. However, it has been well known for some time now that these two techniques are inaccurate for R_{sh} measurements of p-n junctions with X_J below 80nm, due to probe penetration and carrier spilling effects that make the measured electrical junction depth shallower than the metallurgical junction depth (X_J) as determined by SIMS [30, 31]. Other state-of-art electrical characterization techniques such as micro-four-point probe [32], sheet resistance and leakage probe [33], and elastic metal four-point probe [34] have also been shown to be inaccurate for electrical characterization of p-n junctions with sub-15nm X_J [35]. In this work, a new method for accurately determining R_{sh} from electrical measurements and Sentaurus technology computer-aided design (TCAD) simulations is proposed [36, 37]. This R_{sh} extraction method was used to determine the impacts of the OI layers and an LTO capping layer on dopant activation in ultra-shallow junctions [2].

3.4.1 Test Structure Fabrication

Figure 3.8 shows schematically the fabrication process flow to form pairs of $100 \times 100\ \mu m^2$ metal contact pads on the sample (chip) surface, with spacing ranging from 400 μm to 1400 μm. Prior to test structure fabrication, annealed chips received a 120 ^{o}C sulfuric peroxide mixture ($H_2SO_4 : H_2O_2$) bath for 10 minutes, followed by a 1-minute dip in DHF (10:1 H_2O:49% HF) solution to remove the LTO and/or native oxide. Starting from the cleaned USJ sample, a 200nm-thick SiO_2 layer was deposited by low-temperature (350^{o}C) PECVD. This oxide layer serves to protect the USJ sample surface from the photoresist developer solution, which contains TMAH. Then, the sample was spin-coated with 400nm-thick lift-off resist (LOR-3A) and 2 μm-thick g-line resist in a Headway^{TM} photoresist spinner. The LOR-3A film serves as an undercut layer in a bi-layer lift-off process. Then, the bi-layer photoresist stacks were patterned by photolithography. Probing regions were exposed by etching the LTO layer in DHF (10:1 H_2O:49% HF) solution for 6 minutes, then metal films (5nm titanium / 500nm aluminum) were immediately deposited using e-beam evaporation.

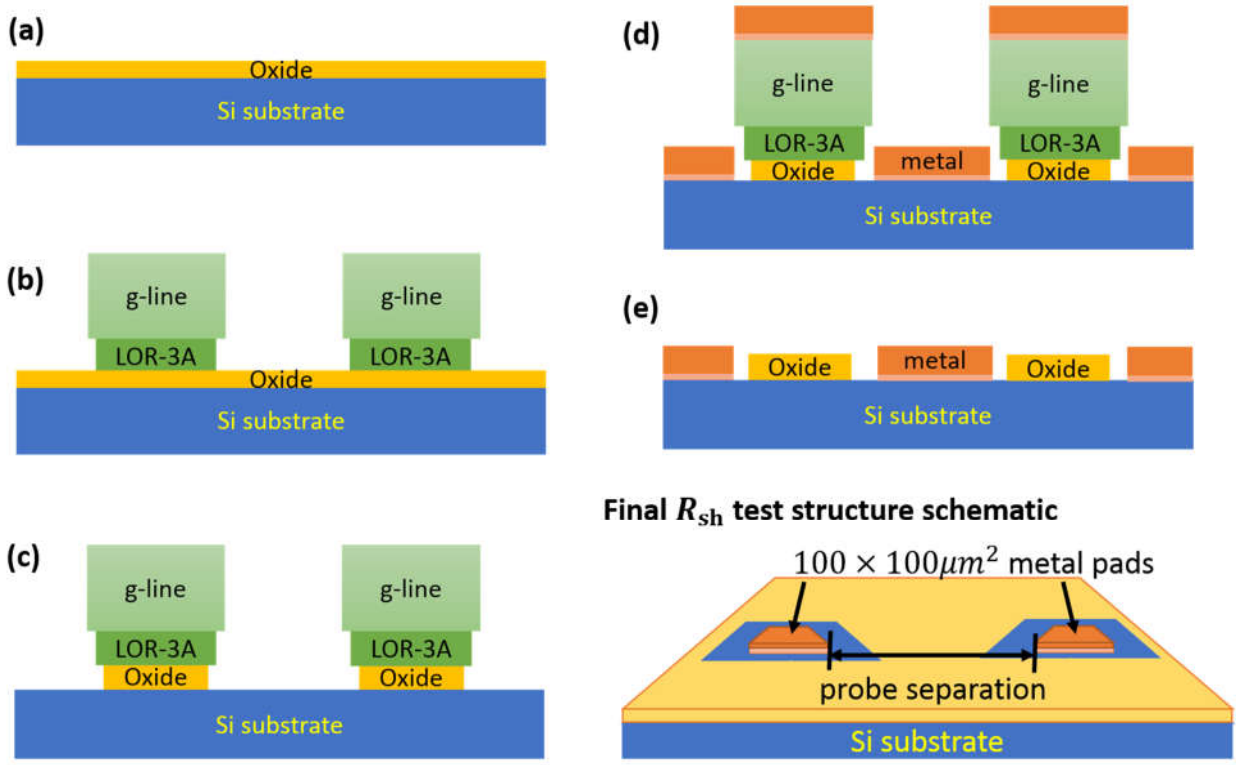

Fig. 3.8. Fabrication process flow for R_{sh} test structure.

3.4.2 R_{sh} Extraction

The R_{sh} extraction method employed in this study is reminiscent of the variable probe-spacing (VPS) measurement method [38]. During a VPS experiment, the spreading resistance on the sample surface is measured for five to seven probe separations. Probe separations range from 20-30 μm to 1000 μm for probe tips of radius 1 μm. To ensure repeatability and accuracy, 20-30 data points are collected for each probe separation. However, a recent study showed VPS is inaccurate for electrical characterization of sub-15nm X_J. It was measured by an atomic force microscope (AFM) that a loading force of 5 g causes a probe penetration of 5 nm [39]. This probe penetration can affect R_{sp} measurements by substrate conduction. In previous studies, it was hypothesized that R_{sh} is directly proportional to R_{sp}. Averaged spreading resistance values are plotted against the natural logarithm of the probe separation. The product of the slope of this plot and π is calculated as R_{sh}, i.e., $R_{\mathrm{sh}} = \pi \times dR_{\mathrm{sp}}/d\ln(\text{separation})$. The impact of this over-simplified R_{sh} extraction method on electrical characterization is demonstrated by the TCAD simulations, as shown in Figure 3.9. Simulated R_{sp} data points were plotted against various probe tip separation values for two heavily-doped p-n junctions, assuming an ideal

step-function profile and 100% electrical activation. It is reasonable to assume negligible metal-semiconductor contact resistance and probe resistance due to the high doping level chosen. Table 3.5 compares TCAD simulated R_{sh} against fitted R_{sh} values. It can be seen that the widely-adopted but over-simplified R_{sh} fitting is inaccurate even for the case when the X_J (100 nm) is much deeper than the probe penetration depth (5 nm).

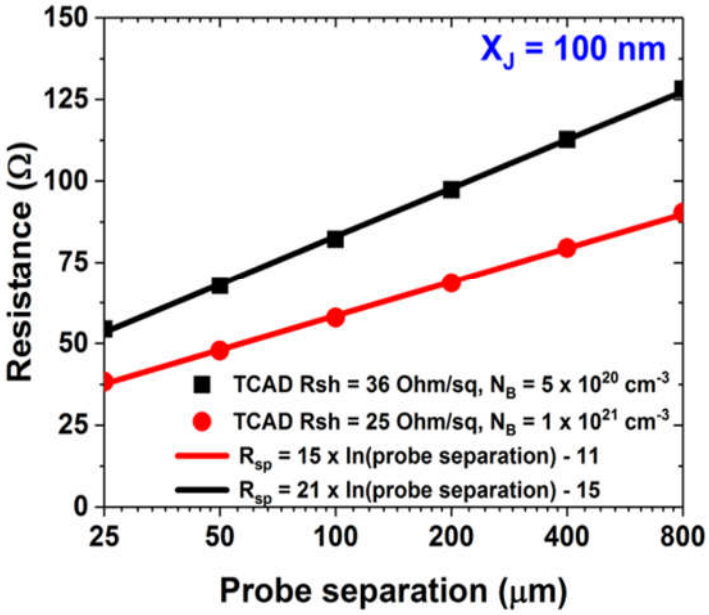

Fig. 3.9. Spreading resistance *vs.* probe tip of radius 1 μm separation for ideal B-doped USJs. The symbols correspond to TCAD simulated data, and the lines are fitted to the data using the least squares method.

N_{Boron} (cm^{-3})	$R_{sh,\ TCAD}$ (Ω/sq)	$R_{sh,\ fit}$ (Ω/sq)	Error (%)
5×10^{20}	35	66	85
1×10^{21}	25	47	88

Table 3.5. Comparison of TCAD simulated R_{sh} *vs.* linear-fitted R_{sh} values. Linear-fitting can result in errors higher than 80%, even when X_J (100 nm) is much deeper than probe penetration (5 nm).

To prevent probe penetration issues, 550 nm thick metal contact pads were deposited onto the USJ sample surface in this work. During the spreading resistance (R_{sp}) measurement, the applied voltage was swept from -10 to 10 mV and the samples were kept at 25 ^{o}C. Ten measurements were taken for each pad separation; averaged R_{sp} values were plotted against the pad separation data. Ohmic contact behavior was observed for

all samples. TCAD simulations were used to determine R_{sh} by fitting experimental results using the least squares method, as illustrated in Figures 3.10-12. It is worth noting that the standard deviation of these measurements is too small to be shown.

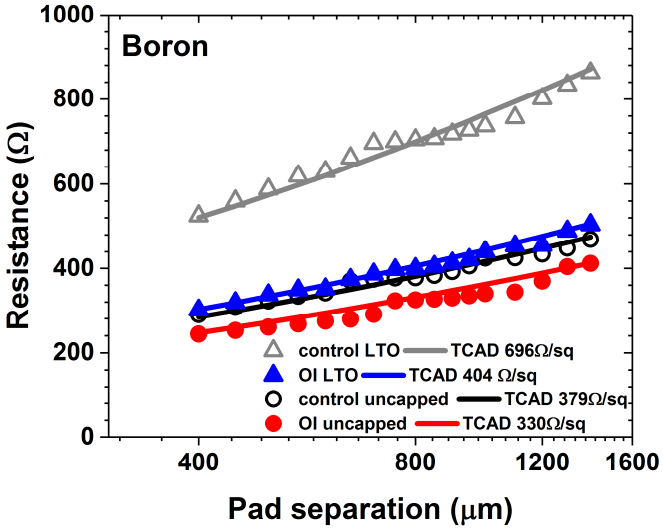

Fig. 3.10. Spreading resistance values *vs.* pad separation data for B-doped samples. Symbols correspond to averaged measurements of R_{sp}, and the lines correspond to the TCAD simulation result fitted to the measurements using the least squares method.

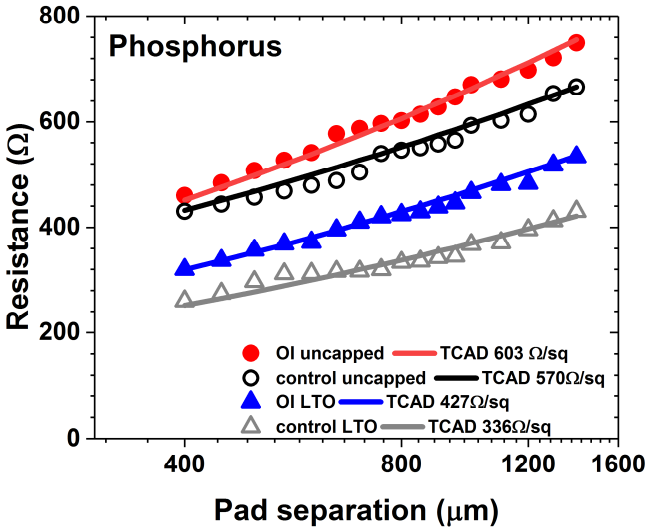

Fig. 3.11. Spreading resistance values *vs.* pad separation data for P-doped samples. Symbols correspond to averaged measurements of R_{sp}, and the lines correspond to the TCAD simulation result fitted to the measurements using the least squares method.

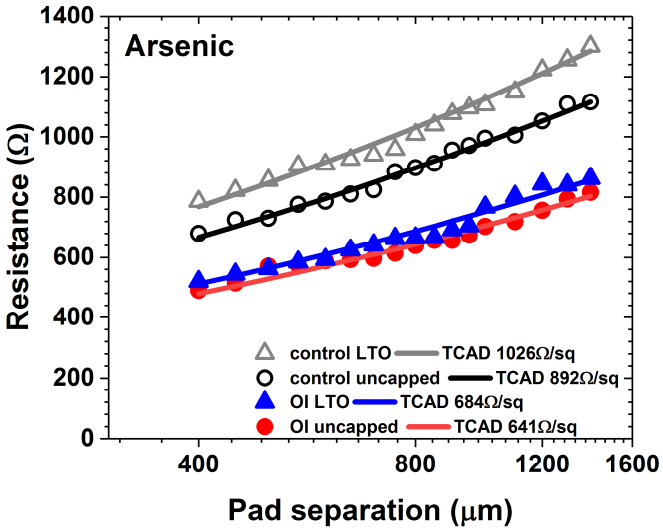

Fig. 3.12. Spreading resistance values *vs.* pad separation data for arsenic-doped samples. Symbols correspond to averaged measurements of R_{sp}, and the lines correspond to the TCAD simulation result fitted to the measurements using the least squares method.

3.4.3 Dopant Activation Analysis

Dopant segregation, clustering, and precipitation limit the electrical dopant activation level, $N_{max,active}$ that can be achieved. In this sub-section, a TCAD simulation methodology based on SIMS dopant depth data was employed to determine the value of $N_{max,active}$ by fitting simulations to measurements, assuming 0% dopant activation at concentration values above $N_{max,active}$. The impacts of the OI layers and an LTO capping layer on dopant activation are determined using this approach.

Figure 3.13 shows the simulated dependence of R_{sh} on $N_{max,active}$ for uncapped and LTO capped B samples. The symbols denote the fitted values of $N_{max,active}$ corresponding to the extracted values of R_{sh} (cf. Figure 3.10). Table 3.6 summarizes X_J, retained B dose, and fitted R_{sh} and $N_{max,active}$ values for B-doped USJs. It can be seen that $N_{max,active}$ values are higher than the plateau level of the diffusion tail, consistent with previous findings [40]. Since it is energetically favorable for interstitial silicon atoms to locate around the OI layers, the diffusion of Si interstitials into the substrate is reduced in the presence of the OI layers and thereby retard the diffusion of B atoms away from the surface. Thanks to the enhanced B dose retention capability during the 1050 °C spike anneal, the OI layers can provide for lower R_{sh} along with reduced X_J.

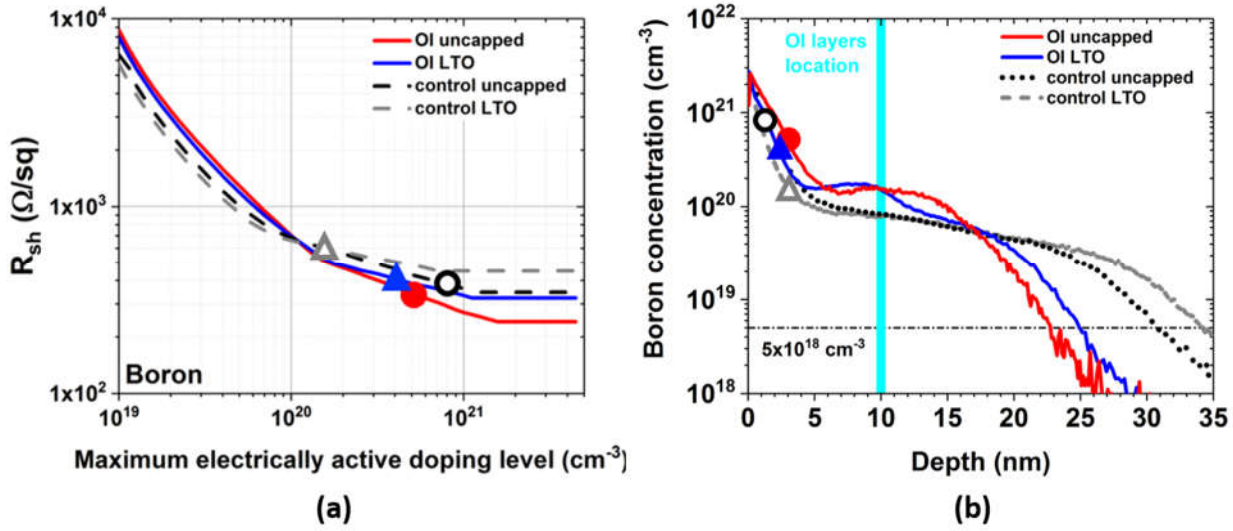

Fig. 3.13. (a) Simulated R_{sh} *vs.* $N_{max,active}$ for B-doped USJ samples, assuming 0% dopant activation at concentration values above $N_{max,active}$. The red solid circle indicates $N_{max,active}$ corresponding to the extracted R_{sh} value (*cf.* Figure 3.10); the black open circle, blue solid triangle, and gray open triangle similarly denote $N_{max,active}$ for uncapped control sample, LTO capped OI sample, and LTO capped control sample, respectively. (b) SIMS B depth profile data.

It can be seen that $N_{max,active}$ values are much lower in capped samples than that for uncapped samples, but this effect is alleviated in OI samples. Previous work showed that the out-diffusion of Si interstitials from the substrate is responsible for the correlation between B deactivation and uphill diffusion. As Si interstitials diffuse toward the surface, they either deactivate B via clustering [41] or cause B dose loss via the kick-out mechanism [8, 27]. In uncapped samples, interstitials also recombine with vacancies at the free silicon substrate surface. This recombination process reduces the contributions of Si interstitials to B dose loss and deactivation. Therefore, the activation level of B atoms, $N_{max,active}$ can be expected to be lower in LTO capped samples. This is consistent with the observed steeper B concentration gradient and higher fitted $N_{max,active}$ values in the case of uncapped samples. However, since the OI layers trap silicon interstitials, B deactivation caused by an LTO cap is much less so for OI samples than for control samples.

B sample	X_J (nm)	Dose ($\times 10^{14}$ cm^{-2})	$N_{max,active}$ ($\times 10^{20}$ cm^{-3})	R_{sh} (Ω/sq)
Uncapped control	31	4.5	8	379
Uncapped OI	22	6.2	5	330
LTO capped control	34	3.6	1.5	696
LTO capped OI	25	4.7	4	404

Table 3.6. Summary of X_J, retained B dose, extracted R_{sh} and $N_{max,active}$ values for B-doped USJs.

Figure 3.14 plots the simulated dependence of R_{sh} on $N_{max,active}$ for uncapped and LTO capped P samples. The symbols denote the fitted values of $N_{max,active}$ corresponding to the extracted values of R_{sh} (cf. Figure 3.11). Table 3.7 summarizes X_J, retained P dose, and fitted R_{sh} and $N_{max,active}$ values for P-doped USJs. It can be seen that P is most vulnerable to dose loss associated with out-diffusion of Si interstitials during RTA by comparing the retained dopant doses in uncapped samples from Tables 3.6, 3.7, 3.8. Due to the impeded diffusion of Si interstitials into the substrate, both uphill diffusion and corresponding P dose loss are enhanced with the OI layers in the uncapped sample. The lower P concentration explains the lower fitted $N_{max,active}$ for uncapped OI sample. Since P prefers to stay in Si over in SiO_2, P dose loss is reduced with an oxide cap.

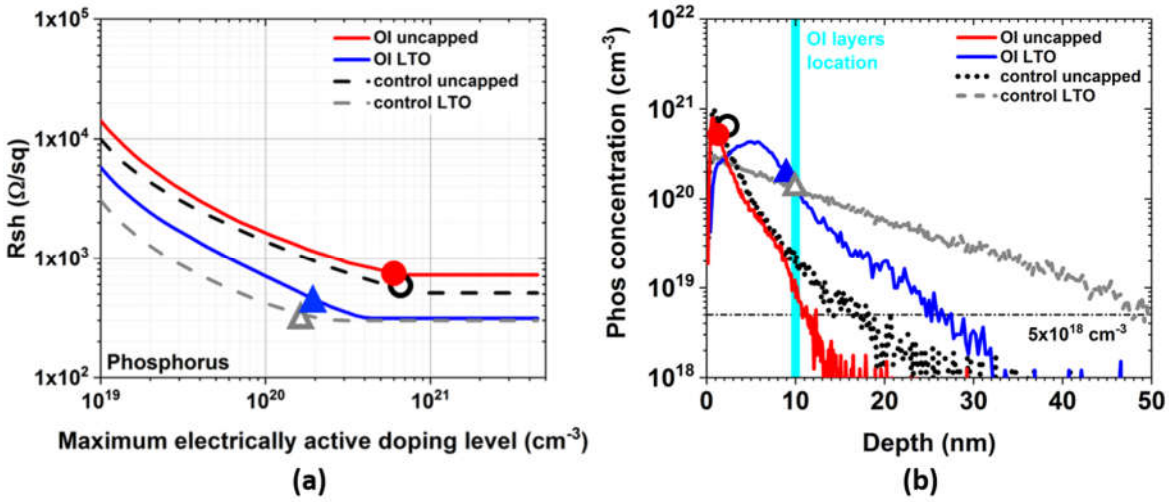

Fig. 3.14 (a) Simulated R_{sh} *vs.* $N_{max,\ active}$ for P-doped USJ samples, assuming 0% dopant activation at concentration values above $N_{max,active}$. The red solid circle indicates $N_{max,active}$ corresponding to the extracted R_{sh} value (*cf.* Figure 3.11); the black open circle, blue solid triangle, and gray open triangle similarly denote $N_{max,active}$ for the uncapped control sample, LTO capped OI sample, and LTO capped control sample, respectively. (b) SIMS P depth profile data.

P sample	X_J (nm)	Dose ($\times 10^{14}$ cm^{-2})	$N_{max,active}$ ($\times 10^{20}$ cm^{-3})	R_{sh} (Ω/sq)
Uncapped control	17	2.2	6	570
Uncapped OI	11	1.6	5	603
LTO capped control	49	3.6	1.7	336
LTO capped OI	26	3.4	2	427

Table 3.7. Summary of X_J, retained B dose, extracted R_{sh} and $N_{max,active}$ values for P-doped USJs.

Figure 3.15 plots the simulated dependence of R_{sh} on $N_{max,active}$ for uncapped and LTO capped As samples. The symbols denote the fitted values of $N_{max,active}$ corresponding to the extracted values of R_{sh} (cf. Figure 3.12). Table 3.8 summarizes X_J, retained As dose, and fitted R_{sh} and $N_{max,active}$ values for As-doped USJs. Thanks to the higher As concentration values, the OI layers can be seen to provide for higher $N_{max,active}$. It can be expected that the free surface serving as a trap of interstitials has lesser impact on As diffusion than for B and P because As diffusion is only 40% interstitial-driven at 1050°C [27]. This can explain why the impacts of an LTO capping layer on X_J and $N_{max,active}$ are much less for As than for B and P. The OI layers are beneficial for improving As USJ conductivity thanks to both higher $N_{max,active}$ and higher retained As dose.

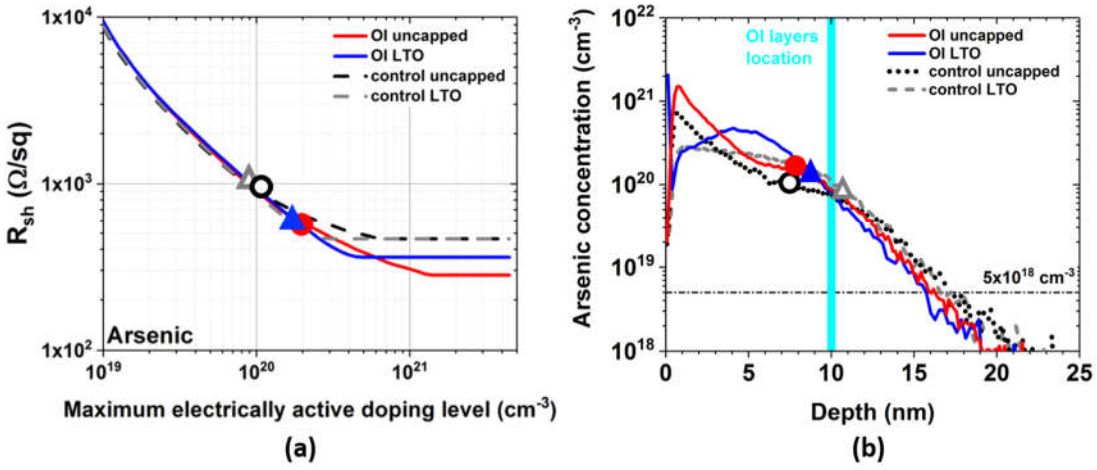

Fig. 3.15. (a) Simulated R_{sh} *vs.* $N_{max,\ active}$ for As-doped USJ samples, assuming 0% dopant activation at concentration values above $N_{max,active}$. The red solid circle indicates $N_{max,active}$ corresponding to the extracted R_{sh} value (*cf.* Figure 3.12); the black open circle, blue solid triangle, and gray open triangle similarly denote $N_{max,active}$ for the uncapped control sample, LTO capped OI sample, and LTO capped control sample, respectively. (b) SIMS As depth profile data.

As sample	X_J (nm)	Dose ($\times 10^{14}$ cm^{-2})	$N_{max,active}$ ($\times 10^{20}$ cm^{-3})	R_{sh} (Ω/sq)
Uncapped control	17	2.4	1	892
Uncapped OI	16	4.0	1.6	641
LTO capped control	17	2.3	0.9	1026
LTO capped OI	16	3.0	1.5	684

Table 3.8. Summary of X_J, retained B dose, extracted R_{sh} and $N_{max,active}$ values for As-doped USJs.

3.4.4 X_J versus R_{sh} Trade-off

Figures 3.16-18 plot extracted R_{sh} values against junction depth (X_J) for B, P, and As-doped USJs, respectively. The OI layers are advantageous to reduce X_J and mitigate R_{sh} increase for B and As doped samples. In the case of P samples, the increase in R_{sh} is less than expected with the OI layers. Table 3.9 summarizes the extracted values of resistivity (*i.e.*, $\rho = R_{sh} \times X_J$) of the heavily doped region. The OI layers are shown to provide for lower ρ values. On the contrary, an LTO capping layer results in higher ρ values, suggesting that an oxide cap does not facilitate low-resistivity USJ formation.

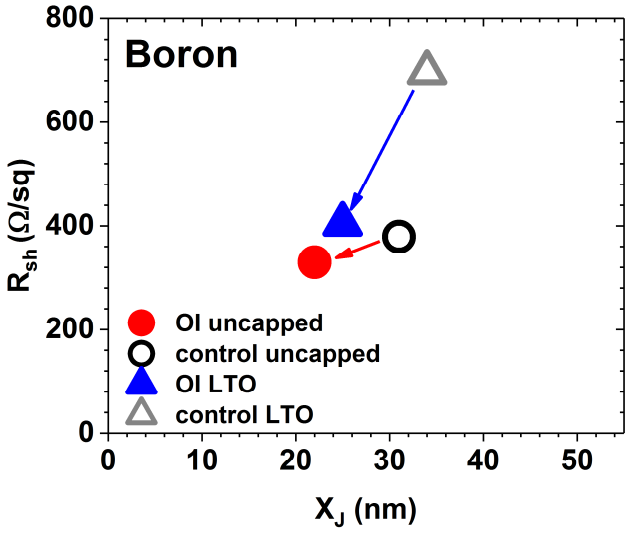

Fig. 3.16. Impacts of the OI layers and LTO capping layer on R_{sh} and X_J, for ultra-shallow junctions formed by B ion implantation and RTA anneal.

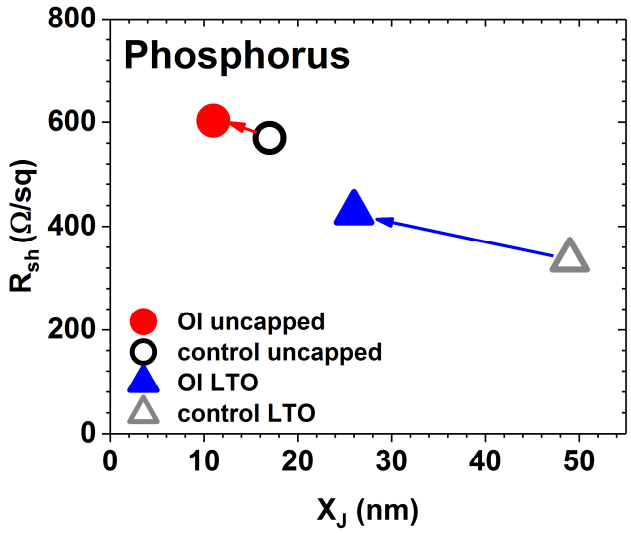

Fig. 3.17. Impacts of the OI layers and LTO capping layer on R_{sh} and X_J, for ultra-shallow junctions formed by P ion implantation and RTA anneal.

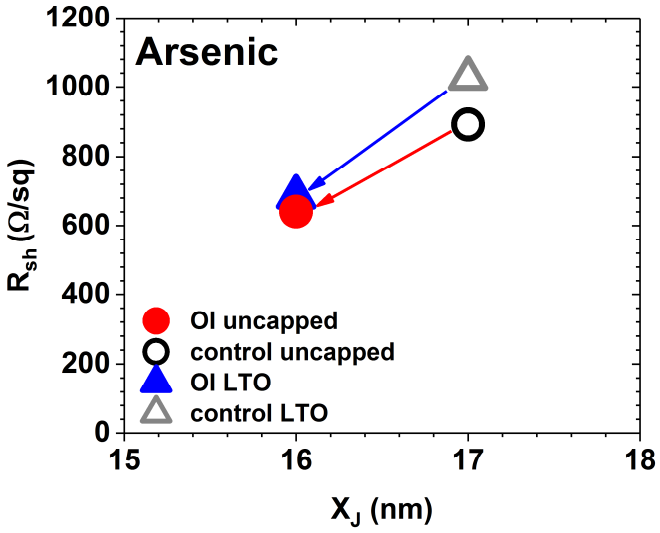

Fig. 3.18. Impacts of the OI layers and LTO capping layer on R_{sh} and X_J, for ultra-shallow junctions formed by As ion implantation and RTA anneal.

$R_{sh} \times X_J$ ($nm \times \Omega/sq$)	B doped	P doped	As doped
Uncapped control	11749	9690	15164
Uncapped OI	7260	6633	10256
LTO capped control	23664	16464	17442
LTO capped OI	10100	11102	10944

Table 3.9. Comparison of $R_{sh} \times X_J$ for B, P, and As doped USJs.

3.5 Summary

OI layers are found to be beneficial for reducing X_J by impeding the diffusion of Si interstitials, whereas neither a SiN_x capping layer nor an LTO capping layer helps to reduce X_J. In addition, the OI layers can mitigate the increase in R_{sh} with X_J scaling due to enhanced dose retention capability during RTA treatment. On the contrary, an LTO capping layer causes an increase in R_{sh} because of lower peak active dopant concentration. This detrimental effect is found to be alleviated in the presence of the OI layers. Its abilities to reduce X_J and to mitigate R_{sh} increase make oxygen insertion technology a promising candidate to facilitate low-resistivity ultra-shallow junction formation for advanced planar CMOSFETs.

3.6 References

[1] X. Zhang *et al.*, "Effects of oxygen-inserted layers on diffusion of boron, phosphorus, and arsenic in silicon for ultra-shallow junction formation," *Journal of Applied Physics,* vol. 123, no. 12, 2018.

[2] X. Zhang *et al.*, "Effects of oxygen-inserted layers and oxide capping layer on dopant activation for the formation of ultrashallow pn junctions in silicon," *Journal of Vacuum Science & Technology B, Nanotechnology and Microelectronics: Materials, Processing, Measurement, and Phenomena,* vol. 36, no. 6, 2018.

[3] B. Davari, R. H. Dennard, and G. G. Shahidi, "CMOS scaling for high performance and low power-the next ten years," *Proceedings of the IEEE,* vol. 83, no. 4, pp. 595-606, 1995, doi: 10.1109/5.371968.

[4] T. Ghani *et al.*, "Scaling challenges and device design requirements for high performance sub-50 nm gate length planar CMOS transistors," in *Symposium on VLSI Technology*, Honolulu, HI, USA, 13-15 June 2000 2000: IEEE, doi: 10.1109/VLSIT.2000.852814.

[5] Y. Taur, C. H. Wann, and D. J. Frank, "25 nm CMOS design considerations," in *IEDM, Technical Digest*, 1998: IEEE, pp. 789-792.

[6] R. H. Dennard, F. H. Gaensslen, V. L. Rideout, E. Bassous, and A. R. LeBlanc, "Design of ion-implanted MOSFET's with very small physical dimensions - IEEE Journals & Magazine," *IEEE Journal of Solid-State Circuits,* vol. 9, no. 5, pp. 256-268, 1974, doi: 10.1109/JSSC.1974.1050511.

[7] S. C. Jain *et al.*, "Transient enhanced diffusion of boron in Si," *Journal of Applied Physics,* vol. 91, no. 11, pp. 8919-8941, 2002, doi: https://doi.org/10.1063/1.1471941.

[8] S. Mirabella, D. D. Salvador, E. Napolitani, E. Bruno, and F. Priolo, "Mechanisms of boron diffusion in silicon and germanium," *Journal of Applied Physics,* vol. 113, no. 3, 2013, doi: https://doi.org/10.1063/1.4763353.

[9] S. M. Hu, P. Fahey, and R. W. Dutton, "On models of phosphorus diffusion in silicon," *Journal of Applied Physics,* vol. 54, no. 12, pp. 6912-6922, 1998, doi: https://doi.org/10.1063/1.331998.

[10] F. F. Morehead and R. F. Lever, "Enhanced "tail" diffusion of phosphorus and boron in silicon: self‐interstitial phenomena," *Applied Physics Letters,* vol. 48, no. 2, pp. 151-153, 1998, doi: https://doi.org/10.1063/1.96980.

[11] P. A. Stolk, H. J. Gossmann, D. J. Eaglesham, D. C. Jacobson, J. M. Poate, and H. S. Luftman, "Trap-limited interstitial diffusion and enhanced boron clustering in silicon," *App. Phys. Let.* vol. 66, no.5, pp.568 - 570, 1994, doi: https://doi.org/10.1063/1.114015.

[12] L. Pelaz *et al.*, "Activation and deactivation of implanted B in Si," *Applied Physics Letters,* vol. 75, no. 5, pp. 662-664, 1999, doi: https://doi.org/10.1063/1.124474.

[13] S. Solmi, E. Landi, and F. Baruffaldi, "High‐concentration boron diffusion in silicon: Simulation of the precipitation phenomena," *J. Appl. Phys.,* vol. 68, no. 7, pp. 3250-3258, 1990, doi: https://doi.org/10.1063/1.346376.

[14] D. Nobili, A. Armigliato, M. Finnetti, and S. Solmi, "Precipitation as the phenomenon responsible for the electrically inactive phosphorus in silicon," *Journal of Applied Physics,* vol. 53, no. 3, pp. 1484-1491, 1998, doi: https://doi.org/10.1063/1.330646.

[15] N. Xu *et al.*, "Extension of Planar Bulk n-Channel MOSFET Scaling With Oxygen Insertion Technology," *IEEE T Electron Dev,* vol. 61, no. 9, pp. 3345-3349, 2014, doi: 10.1109/TED.2014.2342496.

[16] H. Takeuchi *et al.*, "Punch-Through Stop Doping Profile Control via Interstitial Trapping by Oxygen-Insertion Silicon Channel," *IEEE Journal of the Electron Devices Society,* vol. 6, pp. 481-486, 2018, doi: 10.1109/JEDS.2017.2769682.

[17] A. Asenov, "Random dopant induced threshold voltage lowering and fluctuations in sub-0.1 μm MOSFET's: A 3-D "atomistic" simulation study," *IEEE T Electron Dev,* vol. 45, no. 12, pp. 2505-2513, 1998, doi: 10.1109/16.735728.

[18] J. F. Ziegler and J. P. Biersack, "The stopping and range of ions in matter," in *Treatise on heavy-ion science*: Springer, 1985, pp. 93-129.

[19] D. C. Boyd, J. R. Holt, M. Ieong, R. T. Mo, Z. Ren, and G. G. Shahidi, "Ultra-thin body super-steep retrograde well (SSRW) FET devices," USA Patent Appl. 10/710,736, 2006.

[20] H. C. H. Wang, C. C. Wang, C. S. Chang, T. Wang, P. B. Griffin, and C. H. Diaz, "Interface induced uphill diffusion of boron: an effective approach for ultrashallow junction," *IEEE Electron Device Letters,* vol. 22, no. 2, pp. 65 - 67, 2001, doi: 10.1109/55.902833.

[21] R. Duffy *et al.*, "Boron uphill diffusion during ultrashallow junction formation," *Applied Physics Letters,* vol. 82, no. 21, 2003, doi: 1.1578512.

[22] M. Ferri, S. Solmi, A. Parisini, M. Bersani, D. Giubertoni, and M. Barozzi, "Arsenic uphill diffusion during shallow junction formation," *J. Appl. Phys.,* vol. 99, no. 11, p. 113508, 2006.
[23] A. G. Mokhberi, P. B.; Plummer, J. D.; Paton, E.; McCoy, S, "A comparative study of dopant activation in boron, BF_2, arsenic, and phosphorus implanted silicon," *IEEE T Electron Dev,* vol. 49, no. 7, pp. 1183-1191, 2002.
[24] E. Landi, A. Armigliato, S. Solmi, R. Kögler, and E. Wieser, "Electrical activation of boron-implanted silicon during rapid thermal annealing," *Applied Physics A,* vol. 47, no. 4, pp. 359-366, 1988.
[25] H. U. Jager, "An Explanation of Transient-Enhanced Diffusion and Electrical Activation of Boron in Crystalline Silicon during Postimplantation Annealing," *Nucl Instrum Meth B,* vol. 80-1, pp. 679-682, Jun 1993.
[26] H. Wang, C. Wang, C. Chang, T. Wang, P. B. Griffin, and C. H. Diaz, "Interface induced uphill diffusion of boron: An effective approach for ultrashallow junction," *IEEE Electr. Device. L.,* vol. 22, no. 2, pp. 65-67, 2001.
[27] H. Bracht, "Diffusion mechanisms and intrinsic point-defect properties in silicon," *MRS Bull.,* vol. 25, no. 6, pp. 22-27, 2000.
[28] F. Smits, "Measurement of sheet resistivities with the four-point probe," *Bell System Technical Journal,* vol. 37, no. 3, pp. 711-718, 1958.
[29] P. Eyben, M. Xu, N. Duhayon, T. Clarysse, S. Callewaert, and W. Vandervorst, "Scanning spreading resistance microscopy and spectroscopy for routine and quantitative two-dimensional carrier profiling," *Journal of Vacuum Science & Technology B: Microelectronics and Nanometer Structures Processing, Measurement, and Phenomena,* vol. 20, no. 1, pp. 471-478, 2002.
[30] H. S.M., "Between carrier distributions and dopant atomic distribution in beveled silicon substrates," *Journal of Applied Physics,* vol. 53, no. 3, pp. 1499-1510, 1982, doi: https://doi.org/10.1063/1.330648.
[31] T. Clarysse, P. Eyben, N. Duhayon, M. W. Xu, and W. Vandervorst, "Carrier spilling revisited: On-bevel junction behavior of different electrical depth profiling techniques," *J Vac Sci Technol B,* vol. 21, no. 2, pp. 729-736, 2003, doi: 1.1547723.
[32] D. H. Petersen *et al.*, "Review of electrical characterization of ultra-shallow junctions with micro four-point probes," *J. Vac. Sci. Technol. B,* vol. 28, no. 1, pp. C1C27-C1C33, 2010, doi: 10.1116/1.3224898.
[33] V. N. Faifer, M. I. Current, D. Schroder, T. Clarysse, and W. Vandervorst, "Junction photovoltage (JPV) techniques for ultra-shallow junction characterization," *ECS Transactions,* vol. 11, no. 3, pp. 135-147, 2007.
[34] M. Benjamin, R. Hillard, and J. Borland, "Ultra-shallow junction (USJ) sheet resistance measurements with a non-penetrating four point probe," *Nucl. Instrum. Meth. B,* vol. 237, no. 1-2, pp. 351-355, 2005.
[35] T. H. Clarysse *et al.*, "Accurate Sheet Resistance Measurement on Ultra-Shallow Profiles," *MRS Online Proc. Lib. Arc.,* vol. 912, 2011, doi: 10.1557/proc-0912-c05-07.
[36] D. A. Antoniadis, A. G. Gonzalez, and R. W. Dutton, "Boron in Near‐Intrinsic< 100> and< 111> Silicon under Inert and Oxidizing Ambients—Diffusion and Segregation," *J Electrochem Soc,* vol. 125, no. 5, pp. 813-819, 2019, doi: 10.1149/1.2131554.
[37] Synopsys, *Sentaurus Device User Guide, Version G*. Mountain View, CA, 2017.

[38] T. Clarysse, P. Eyben, T. Hantschel, and W. Vandervorst, "Towards sub-10 nm carrier profiling with spreading resistance techniques," *Mat. Sci. Semicon. Proc.,* vol. 4, no. 1-3, pp. 61-66, 2001.
[39] 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.
[40] S. Jain *et al.*, "Metastable boron active concentrations in Si using flash assisted solid phase epitaxy," *J. Appl. Phys.,* vol. 96, no. 12, pp. 7357-7360, 2004.
[41] G. Mannino, V. Privitera, S. Solmi, and N. E. B. Cowern, "Issues on boron electrical activation in silicon: Experiments on boron clusters and shallow junctions formation," *Nucl Instrum Meth B,* vol. 186, pp. 246-255, 2002.

Chapter 4

FinFET Channel Profile Optimization and Performance Enhancement with Oxygen Insertion Technology

In this chapter, the benefits of a super-steep retrograde (SSR) fin channel profile, which can be achieved using OI technology, are quantified using 3-D technology computer-aided design (TCAD) simulations targeting 7/8-nm low-power applications [1]. Besides, the electrostatic benefits of using a silicon-on-insulator (SOI) substrate versus the SSR FinFET technology are investigated via Sentaurus Device TCAD [2]. A calibrated compact model is then used to estimate the 6-transistor static RAM (6T-SRAM) cell performance and yield. Due to enhanced transistor performance and improved robustness against systematic and random sources of variations, SSR FinFET technology is shown to be a cheaper alternative to the SOI substrate for achieving similarly low minimum cell operating voltage ($V_{DD,min}$) for 6T-SRAM bit cells. Both SSR and SOI FinFET technologies are projected to provide for up to 100 mV reduction in $V_{DD,min}$, to facilitate voltage scaling to below 0.50 V.

4.1 Introduction

To maintain good gate control of the electrical potential in the channel region, *i.e.,* good electrostatic integrity, it is necessary to adopt the multi-gate structures for MOSFET with gate lengths below 25 nm [3]. As a result, a three-dimensional (fin-shaped) channel structure straddled on three sides by the gate electrode, known as either the "tri-gate" or "FinFET" design, has been widely adopted by the leading semiconductor companies (namely Intel Corporation [4-6], Samsung Electronics [7], and Taiwan Semiconductor Manufacturing Company [8]). It was first demonstrated at the University of California, Berkeley that the FinFET design can be scaled to sub-25 nm gate lengths [9].

With advancements in CMOS technology, the widening performance gap between dynamic random access memory (DRAM) and processor demands to increase the capacity of on-chip (monolithically integrated with the processer) cache memory (static RAM, or

SRAM), as illustrated in Figure 4.1. A key challenge to achieving necessarily high yield for large SRAM arrays is the increasing threshold voltage (V_{th}) variations due to process-induced variations as L_{gate} scales. In addition to achieving higher device performance (higher transistor ON-state current for a given operating voltage or lower operating voltage for the same transistor ON-state current), FinFET technology can also mitigate the short-channel effects (SCE) and drain-induced barrier lowering (DIBL) for reduced performance sensitivity to process-induced variations, to overcome this challenge.

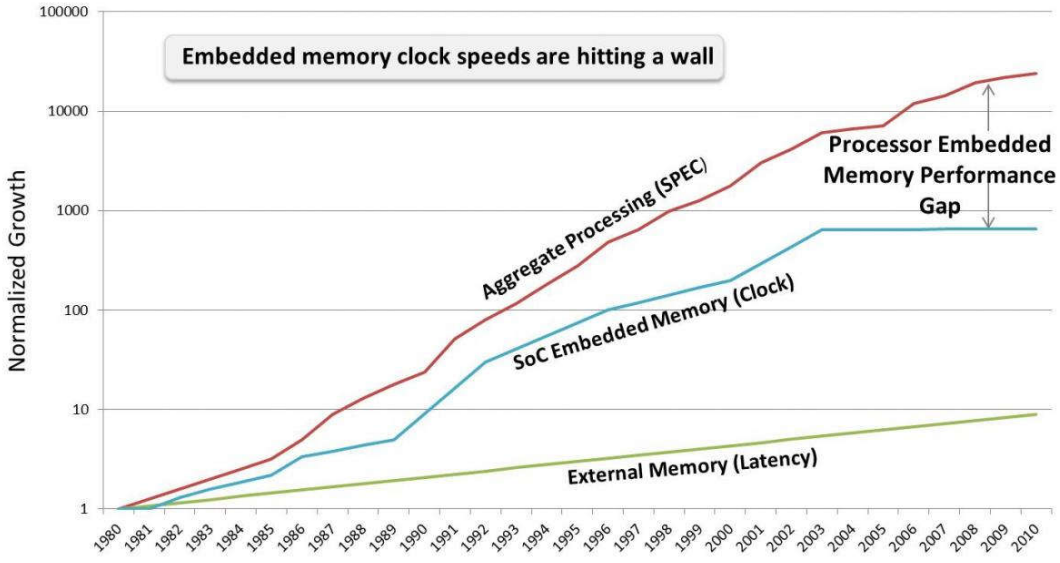

Fig. 4.1. Trend in the performance gap between memory and processor [10].

SOI FinFET technology is the ideal candidate for low-power FinFET technology as the buried oxide (BOX) layer can eliminate the OFF-state leakage current (I_{off}) effectively [11]. The economic downside of the SOI technology is the higher cost of an SOI wafer relative to a conventional bulk-Si wafer. In a bulk-Si FinFET technology, high doping, punch-through stopper (PTS), is necessary at the base of the fins to suppress OFF-state leakage currents. However, a conventional doping process usually results in fin doping on the order of $5 \times 10^{17}\ cm^{-3}$, which can degrade the ON-state current (I_{on}) significantly due to Columbic scattering. As a result, bulk-Si FinFETs have lower I_{on}/I_{off} ratios as compared with SOI FinFETs [12]. The heavily doped channel also results in larger random dopant fluctuation (RDF) induced V_{th} variations (σV_{th}), which is proportional to $1/\sqrt{WL}$ (W: channel width; L: channel length) [13]. In the previous chapter, OI technology is demonstrated to suppress the diffusion of dopants effectively thanks to the impeded diffusion of interstitial Si atoms. Previous work has shown that the OI layers can provide for super-steep retrograde (SSR) doping profiles, *i.e.,* high doping at the base region and light doping at the channel region [14]. This SSR profile formed with the OI layers can

overcome the challenges faced by the conventional bulk-Si FinFET technology, achieving higher I_{on}/I_{off} ratios by reducing Columbic scattering.

In this work, the benefits of an SSR channel doping profile that can be achieved with the OI layers were quantified via 3-D TCAD simulations, targeting the 7/8-nm technology node. The optimized device performance parameters of SSR FinFETs were compared against that for conventional bulk-Si (control) FinFETs and SOI FinFETs. A calibrated compact model was then used to estimate the six-transistor (6T) SRAM performance and yield using the cell sigma method [15].

4.2 Device Simulation and Design Optimization

4.2.1 FinFET Structure

Figure 4.2 schematically illustrates the 3-D FinFET structures simulated via 3-D TCAD simulations in this work. Device design parameter values were chose based on the 7/8-nm low-power technology specifications from the 2013 International Technology Roadmap for Semiconductors [16]. The gate length (L_{gate}) is 15 nm corresponding to the 7/8-nm node. The equivalent oxide thickness is 0.64 nm. The fin height (H_{Si}) is 40 nm, the fin width (W_{Si}) is 8 nm, so the fin aspect ratio is 5. The effective channel width, W_{eff} (*i.e.,* peripheral length of the silicon fin) is 88 nm ($W_{eff} = 2 \times H_{Si} + W_{Si}$). The fin pitch is 30 nm based on Intel 22-nm [5] and 14-nm [6] FinFET technology. To target low-power applications, the gate work function is assumed to be tunable to achieve an OFF-state leakage current specification (I_{off}) of 30 pA/μm, which is consistent with TSMC 16-nm FinFET technology. In this work, current is normalized against W_{eff}. To prevent fringing field effects, the top corners of the fin are rounded (1-nm radius of curvature) for reduced gate leakage, similar as the Intel 14-nm FinFET technology design [6]. In the case of bulk-Si FinFET, the shallow trench isolation (STI) region is 50 nm thick, whereas the buried oxide (BOX) layer in SOI FinFETs is 20 nm thick. To reduce S/D parasitic resistance, the simulated FinFET structures each comprise heavily doped S/D regions formed by selective epitaxial growth (SEG) [17]. In this work, the S/D junctions are assumed to have a Gaussian doping profile with 2-nm/dec gradient and peak concentrations of 2×10^{20} cm^{-3} [18]. In n-channel FinFETs (nFETs), the SEG S/D regions consist of phosphorus-doped silicon. In p-channel FinFETs (pFETs), the SEG S/D regions consist of boron-doped silicon-germanium (SiGe) with 50% germanium concentration, with parameter values based on [19]. As transistor gate length scales, metal/semiconductor ohmic contact resistances become increasingly important to model device performance correctly. In this work, ohmic contacts are assumed to contact only the top surfaces of SEG S/D regions, with specific contact resistivity of $3 \times 10^{-9}\ \Omega \cdot cm^2$. Table 4. 1 summarizes the nominal values of the various design parameters for the simulated FinFETs.

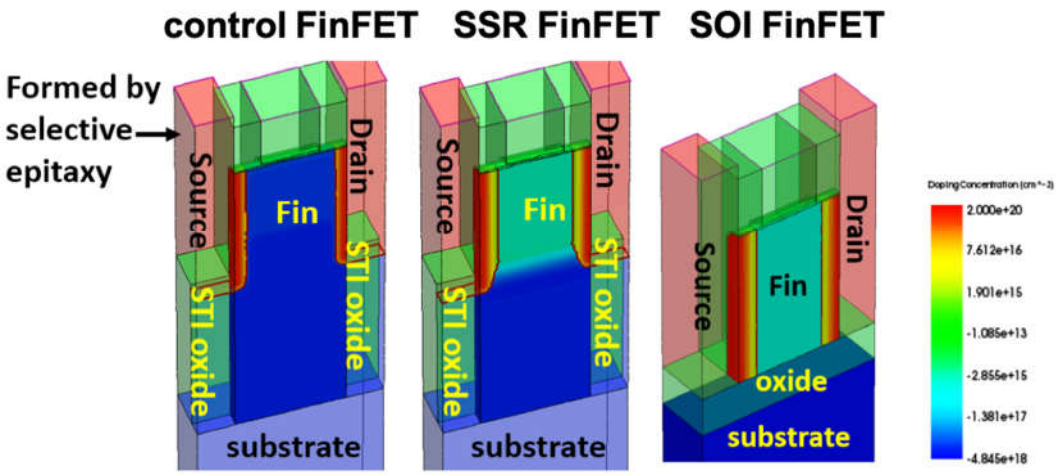

Fig. 4.2. Simulated 3-D n-channel FinFET structures. The net dopant concentration is represented in color using a hyperbolic arcsine scale.

	Control FinFETs		SSR FinFETs		SOI FinFETs	
	n-channel	p-channel	n-channel	p-channel	n-channel	p-channel
L_{gate} (nm)	15	15	15	15	15	15
EOT (nm)	0.64	0.64	0.64	0.64	0.64	0.64
I_{off} (pA/μm)	30	30	30	30	30	30
H_{Si} (nm)	40	40	40	40	40	40
W_{Si} (nm)	8	8	8	8	8	8
N_{fin} (cm^{-3})	5×10^{17}	5×10^{17}	1×10^{15}	1×10^{15}	1×10^{15}	1×10^{15}
$N_{fin,peak}$ (cm^{-3})	2.5×10^{18}	2.5×10^{18}	5×10^{18}	5×10^{18}	1×10^{15}	1×10^{15}
$N_{substrate}$ (cm^{-3})	2.5×10^{18}	2.5×10^{18}	2.5×10^{18}	2.5×10^{18}	1×10^{15}	1×10^{15}
S/D doping gradient (nm/dec)	2	2	2	2	2	2
N_{SD} (cm^{-3})	2×10^{20}	2×10^{20}	2×10^{20}	2×10^{20}	2×10^{20}	2×10^{20}
PTS doping gradient (nm/dec)	> 40	> 40	3.3	6.9	N/A	N/A

Table 4.1. Bulk-Si and SOI FinFETs design: nominal parameter values.

In SOI FinFETs, the fin channel region has a constant doping level of 1×10^{15} cm^{-3}. Since the OI layers are beneficial for suppressing the diffusion of dopants upward into the channel region of CMOS transistors during the front-end-of-line manufacturing process, super-steep retrograde (SSR) fin doping profiles are assumed herein for the bulk-Si FinFETs with the OI layers inserted: the fin channel doping increases from 1×10^{15} cm^{-3} at the fin top to 5×10^{18} cm^{-3} at the punch-through stopper (PTS) layer. Consistent with previous findings [20], the n-channel SSR FinFET was assumed to have a doping gradient of 3.3 nm/dec, and the p-channel SSR FinFET was assumed to have a doping gradient of 6.9 nm/dec. Figure 4.3 provides the cross-sectional view of the simulated n-channel SSR FinFET.

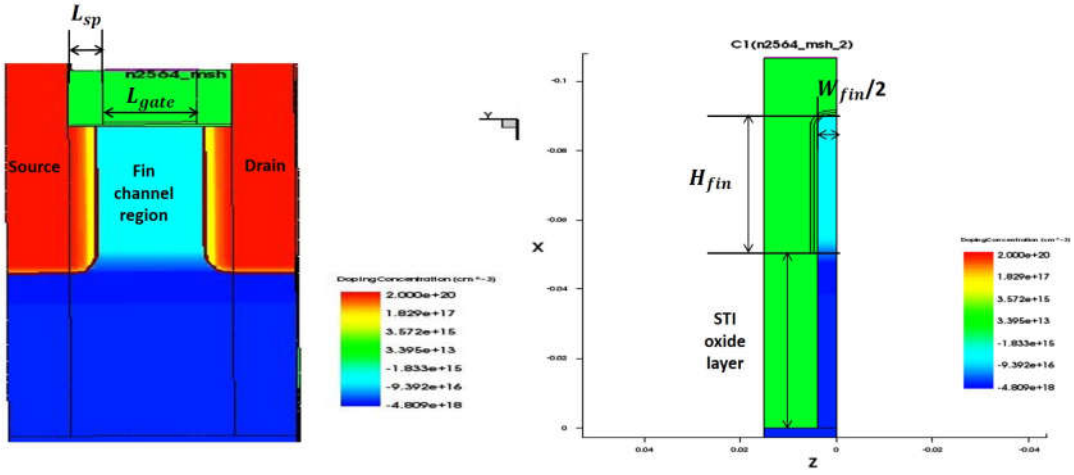

Fig. 4.3. Cross-sectional views of the n-channel SSR FinFET structure. The net dopant concentration is represented in color using a hyperbolic arcsine scale. The fin aspect ratio is $\frac{H_{\text{fin}}}{W_{\text{fin}}} = 5$; the fin shape is rectangular as in Intel's 14-nm FinFET technology [6]; the fin corner radius of curvature is 1 nm.

FinFET performance was simulated using the TCAD software package Sentaurus Device [21], using the drift-diffusion transport model [22] calibrated to ballistic Monte-Carlo simulations, the Philips unified model for carrier mobility, bandgap narrowing model, density gradient quantization model, and nonlocal-path trap-assisted tunneling model [23]. The fin sidewall surfaces (along which the transistor current flows) are assumed to be along {110} crystallographic planes, with transistor current flow in a <110> direction. To boost transistor ON-state current, 2 GPa (tensile) uniaxial stress is assumed for n-channel devices, whereas -2 GPa (compressive) uniaxial stress is assumed for p- channel devices.

4.2.2 Design Optimization Methodology

Effective channel length (L_{eff}) is defined as the distance between the points in the channel where the source/drain dopant concentration drops to 2×10^{19} cm^{-3} [24], and was tuned separately for bulk-Si FinFETs and SOI FinFETs to maximize the ON-state current $I_{d,sat}$

while meeting the same OFF-state leakage specification (I_{off} = 30 pA/μm). In the case of SSR FinFETs, the peak location of the punch-through stopper (PTS) doping profile ($X_{fin,peak}$) was optimized to improve the device performance.

***L_{eff}* Optimization** Previous work showed that tuning L_{eff} could adjust the tradeoff between series resistances and short-channel effect (SCE) [24]. In practice, L_{eff} can be tuned by adjusting the gate-sidewall spacer length (L_{sp}) or the source/drain doping gradient. The source/drain gradient was made steep (2 nm/dec) for low parasitic source/drain series resistances, hence L_{eff} is tuned by adjusting L_{sp} for each device structure to maximize $I_{d,sat}$.

Figure 4.4 shows the simulated dependence of $I_{d,sat}$ on L_{eff} for n-channel FinFETs (nFETs) and p-channel FinFETs (pFETs). The optimal values of L_{eff} are 25 nm for the n-channel control FinFET, 26 nm for the n-channel SSR FinFET and 24 nm for the n-channel SOI FinFET. Since the fin channel is lightly doped in the SSR FinFET, it requires a larger L_{eff} value to suppress SCE effectively.

Due to smaller band gap energy of the $Si_{0.5}Ge_{0.5}$ SEG S/D regions, p-channel FinFETs have larger gate-induced drain leakage (GIDL) due to band-to-band tunneling, which can dominate OFF-state leakage current. Therefore, larger L_{eff} values are required to suppress GIDL effectively for p-channel FinFETs. The optimal values of L_{eff} are 27 nm for the p-channel control FinFET, 28 nm for the n-channel SSR FinFET and 27 nm for the n-channel SOI FinFET. Figure 4.5 shows the BTBT rate distributions within the p-channel bulk-Si FinFETs in the off state, calculated using Kane's model [23].

Figure 4.6 shows the optimized net dopant concentration profiles along the channel region, from the source region to the drain region, and the optimized fin channel doping profiles, for each of the optimized bulk-Si FinFETs. The buried oxide (BOX) layer in SOI FinFETs can effectively eliminate subfin leakage current. Therefore, SOI FinFETs have smaller optimal values of L_{eff}: the optimal values of L_{eff} are 24 nm for the n-channel SOI FinFET and 27 nm for the p-channel SOI FinFET.

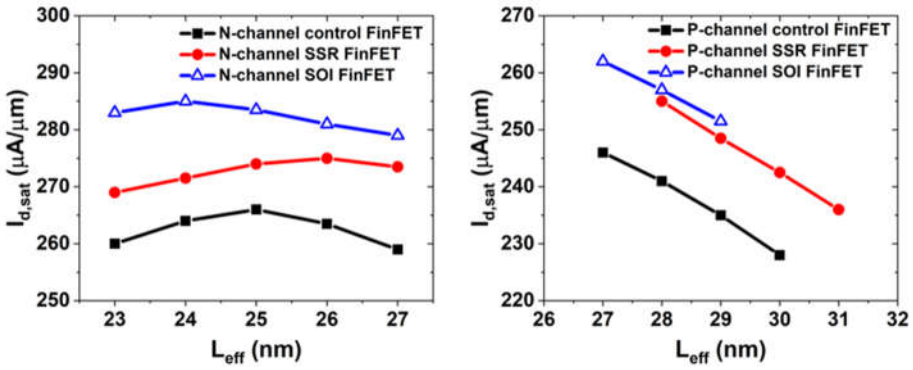

Fig. 4.4. On-state current ($I_{d,sat}$) normalized to W_{eff} *vs.* effective channel length (L_{eff}). Longer L_{eff} values are required to suppress BTBT-induced GIDL effectively for p-channel

FETs.

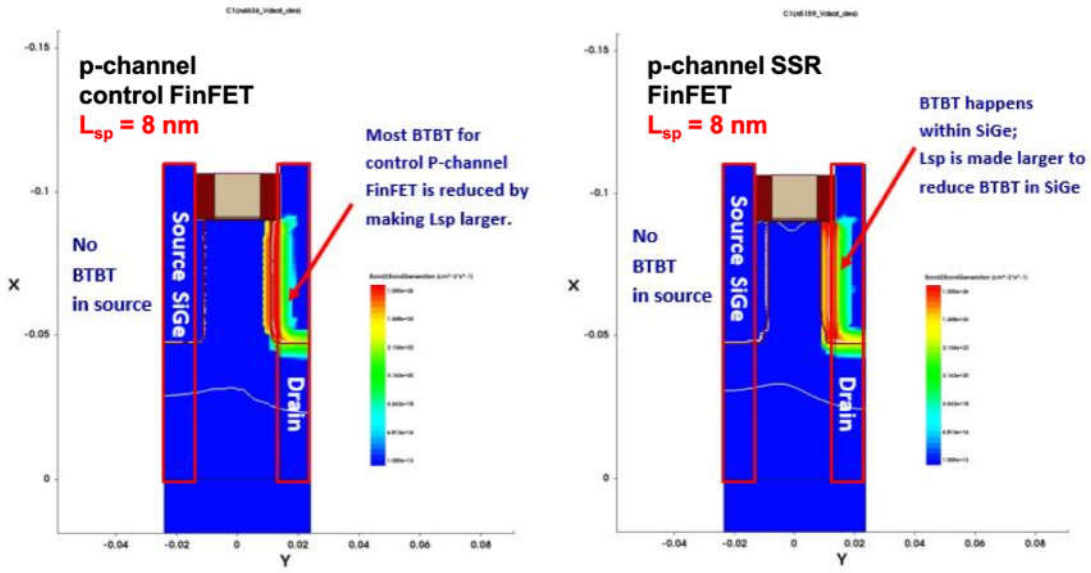

Fig. 4.5. Band-to-band-tunneling rate contour plots for p-channel FinFETs in the off state ($V_{GS} = 0V,\ V_{DS} = V_{DD}$)

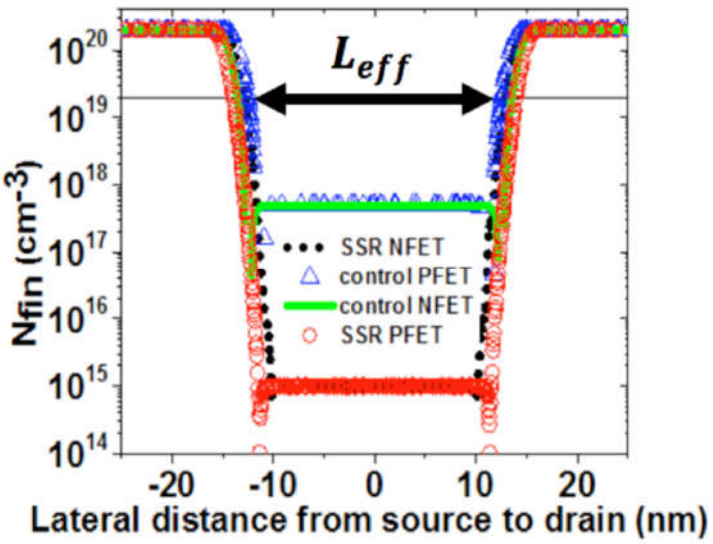

Fig. 4.6. Net dopant concentration profiles along the channel direction, from the source region to the drain region, for the optimized control FinFETs and SSR FinFETs.

Punch-through stopper (PTS) Optimization The previous chapter showed that the oxygen layers within silicon effectively suppress dopant diffusion, causing dopant atoms to pile up; thus, inserting partial oxygen monolayers at a depth corresponding to the base of the silicon fin would facilitate the formation of a super-steep retrograde (SSR) fin doping profile. Figure 4.7 compares the optimized fin doping profiles for control FinFETs and

SSR FinFETs. Since the dopant diffusion blocking effect is greater for boron than that for n-type dopants [20], the retrograde doping gradient is 3.3 nm/dec and 6.9 nm/dec for n-channel SSR and p-channel SSR FinFETs, respectively. To maximize $I_{d,sat}$, both the peak PTS dopant concentration ($N_{fin,peak}$) and the location of the peak ($X_{fin,peak}$) were separately optimized for n-channel SSR and p-channel SSR FinFETs. The optimal value of $X_{fin,peak}$ was 46 nm and the optimal value of $N_{fin,peak}$ was 5×10^{18} cm^{-3} for both n-channel SSR and p-channel SSR FinFETs.

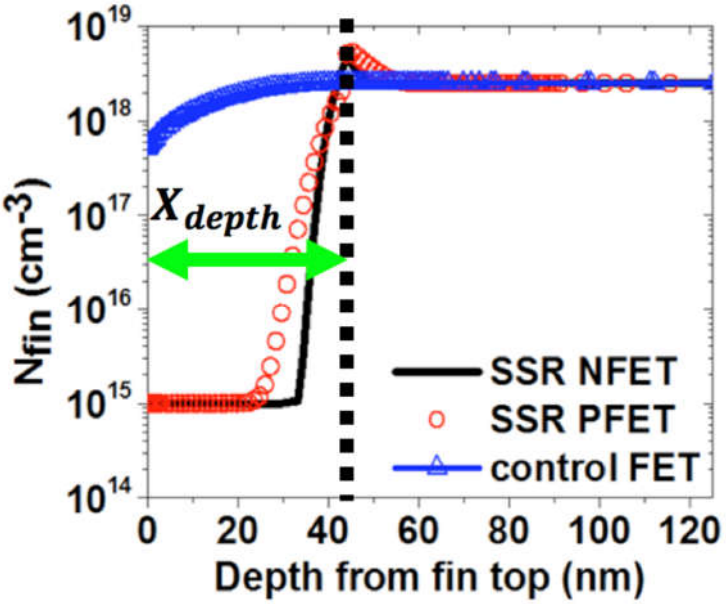

Fig. 4.7. Optimized fin doping depth profiles for the optimized control FinFETs and SSR FinFETs.

Table 4.2 summarizes the key performance parameters for the optimized FinFET designs. Threshold voltage (V_{th}) is extracted based on a constant current criterion of $100\,nA \times \frac{W_{\text{eff}}}{L_{\text{gate}}}$. According to ITRS 2013 specifications, the operating voltage (V_{DD}) should be 0.80 V for 7/8-nm low-power technology node [16]. Compared with control FinFETs, the lightly doped fin channel regions in SSR FinFETs can provide for higher carrier mobility. Therefore, SSR FinFETs provide for 3.6% and 3.8% improvement in $I_{d,sat}$ for nFETs and pFETs. The benefits of adopting an SSR channel profile were greater in the linear regime ($V_{GS} = 0.80\ V,\ V_{DS} = 50\ mV$): SSR FinFETs provide for 6.7% and 6% improvement in $I_{d,sat}$ for nFETs and pFETs. SOI FinFETs have smaller optimal L_{eff} values as compared with SSR FinFETs because the BOX layer more effectively eliminates sub-fin leakage current. The steeper subthreshold swing and lower drain-induced barrier lowering (DIBL = $|(V_{t,sat} - V_{t,lin})/(0.80 - 0.05)|$) suggest that SOI FinFETs have superior electrostatic integrity. Thanks to the reduced S/D parasitic resistances (smaller L_{eff}) and superior electrostatic integrity, SOI FinFETs can provide for 3.8% and 2.8% improvement

in $I_{d,sat}$ for nFETs and pFETs as compared with SSR FinFETs. The performance improvement for pFETs with SOI technology is less than for nFETs due to the smaller reduction in L_{eff}. This is because GIDL current becomes dominant in OFF-state leakage as a consequence of the smaller band gap energy of the SEG $Si_{0.5}Ge_{0.5}$ S/D regions for pFETs; thus, a larger L_{eff} value is required to meet the OFF-state leakage current specification.

	Control FinFETs		SSR FinFETs		SOI FinFETs	
	n-channel	p-channel	n-channel	p-channel	n-channel	p-channel
V_{DD} (V)	0.8	0.8	0.8	0.8	0.8	0.8
$V_{DS,lin}$ (V)	0.05	0.05	0.05	0.05	0.05	0.05
L_{eff} (nm)	25	27	26	28	24	27
Spacer length L_{sp} (nm)	7	8	7.5	8.5	6.5	8
Work function (eV)	4.56	4.67	4.57	4.64	4.58	4.64
$V_{t,sat}$ (V)	0.254	-0.273	0.250	-0.258	0.248	-0.253
$V_{t,lin}$ (V)	0.288	-0.302	0.279	-0.281	0.278	-0.276
$I_{d,sat}$ (μA/μm)	266	246	275	255	285	262
$I_{d,lin}$ (μA/μm)	63	68	67	72	75	70
SSwing (mV/dec)	68	66	69	67	68	66
DIBL (mV/V)	45	39	40	31	40	30

Table 4.2. Summary of key performance parameters for the optimized FinFET designs.

4.2.3 Compact Model Calibration

In this section, an analytical compact model for transistor current as a function of applied voltages is employed to estimate 6T-SRAM cell performance and yield, following the methodology established in [15]. The compact model is based on the short-channel MOSFET I-V equations, accounting for channel length modulation (CLM), velocity saturation, and bulk charge effects. Due to the lack of predictability of GIDL current, 3-D device simulation results without GIDL were used to calibrate the compact model, as

shown in Figures 4.8 – 10.

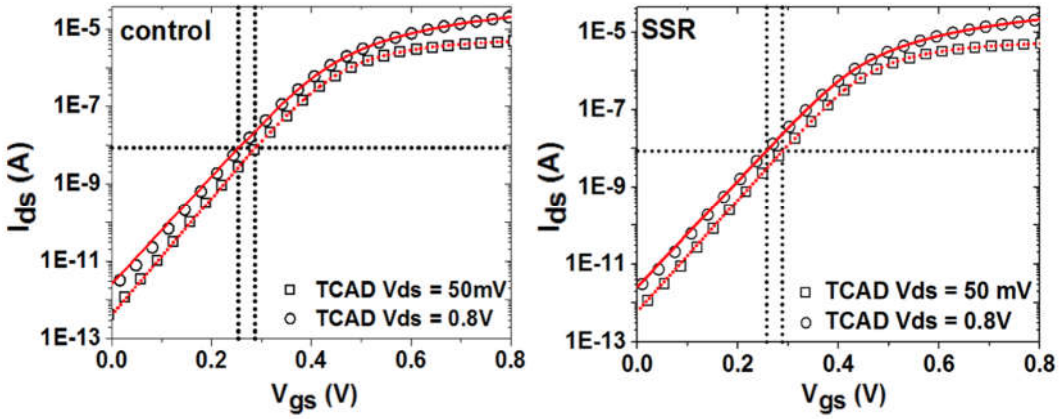

Fig. 4.8. Comparison of the calibrated compact model (lines) and simulated I-V characteristics for n-channel bulk-Si FinFETs.

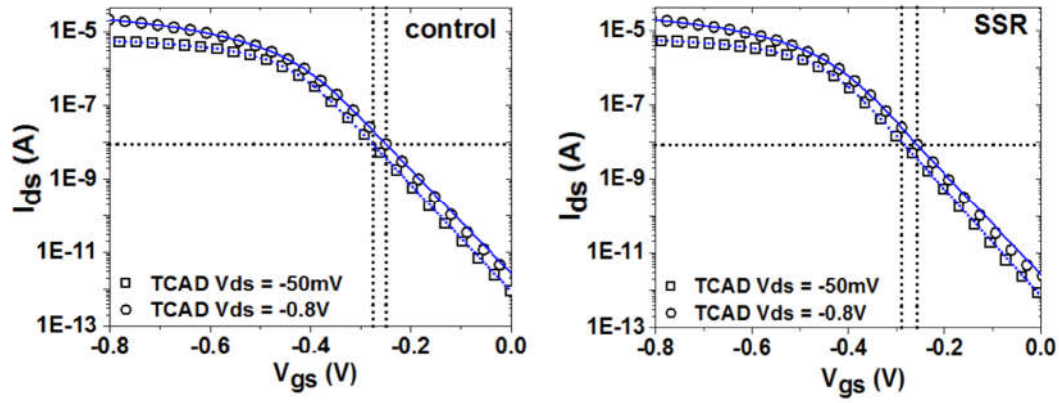

Fig. 4.9. Comparison of the calibrated compact model (lines) and simulated I-V characteristics for p-channel bulk-Si FinFETs.

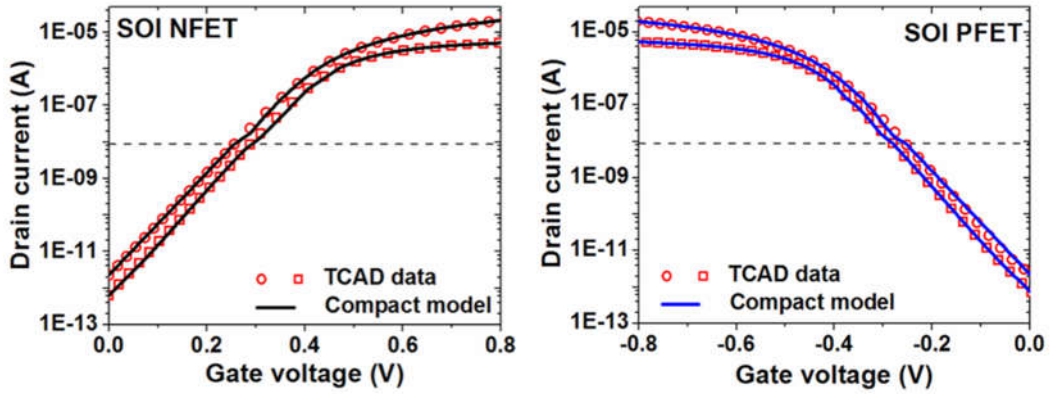

Fig. 4.10. Comparison of the calibrated compact model (lines) and simulated I-V characteristics for SOI FinFETs.

4.3 Variability Study

In this section, the benefits of an SSR fin doping profile which can be achieved with OI technology and SOI technology for improving the immunity of FinFET performance to variations are investigated. Sources of variations can be categorized as either systematic, caused by process variations, or random, caused by intrinsic variations [25, 26]. Process-induced variations in transistor gate length (L_{gate}) and fin width (W_{fin}) are assumed to have Gaussian distributions with $\pm$ 10% variation corresponding to three standard deviations away from the mean (nominal) value. The effects of process-induced variations on transistor threshold voltage (V_{th}) and OFF-state (I_{off}) are shown to be accurately predicted by the compact model. Random sources of variations become dominant and can limit the IC manufacturing yield as transistors are scaled down toward atomic dimensions [27-32]. These random sources of variations include random dopant fluctuations (RDF) [13, 33], and gate work function variation (WFV) [29, 30]. The impact of intrinsic variations is quantified using the noise-like impedance field method (IFM) [33, 34]. Finally, 6T-SRAM cell performance and yield as a function of cell operating voltage are estimated using the calibrated compact model, to quantify the benefits of OI technology and SOI technology, respectively.

4.3.1 Impact of Systematic Variations

Figure 4.11 plots transistor threshold voltage (V_{th}) and OFF-state (I_{off}) *vs.* L_{gate}. These systematic variations are accurately predicted by the calibrated compact model. With L_{gate} decreasing, the saturation threshold voltage ($V_{t,sat}$) and linear threshold voltage ($V_{t,lin}$) both decrease due to the short-channel effect; whereas the OFF-state leakage current (I_{off}) correspondingly increases. Due to the steeper subthreshold swing (SSwing) values, both SOI and SSR FinFETs show slightly greater sensitivity of I_{off} to changes in L_{gate} since $\log(I_{off}) \propto \frac{-V_{th}}{\text{SSwing}}$.

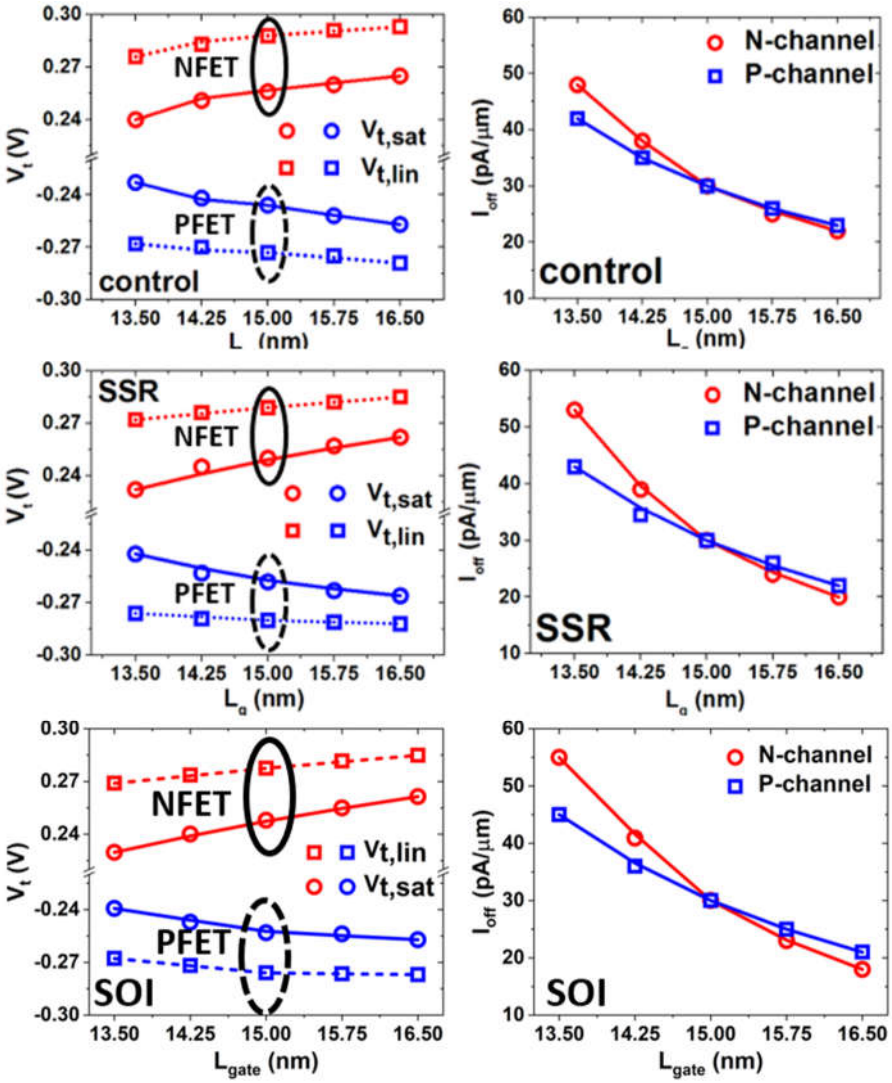

Fig. 4.11. Effects of gate-length (L_{gate}) variation on FinFET threshold voltage (V_{th}) and OFF-state leakage current (I_{off}).

Figure 4.12 plots the dependencies of V_{th} and I_{off} on the fin width (W_{fin}). The calibrated compact model accurately predicts these systematic variations. With W_{fin} scaling, the electrostatic integrity (*i.e.,* gate control) improves, and the quantum confinement effect increases; thus, V_{th} increases and I_{off} decreases.

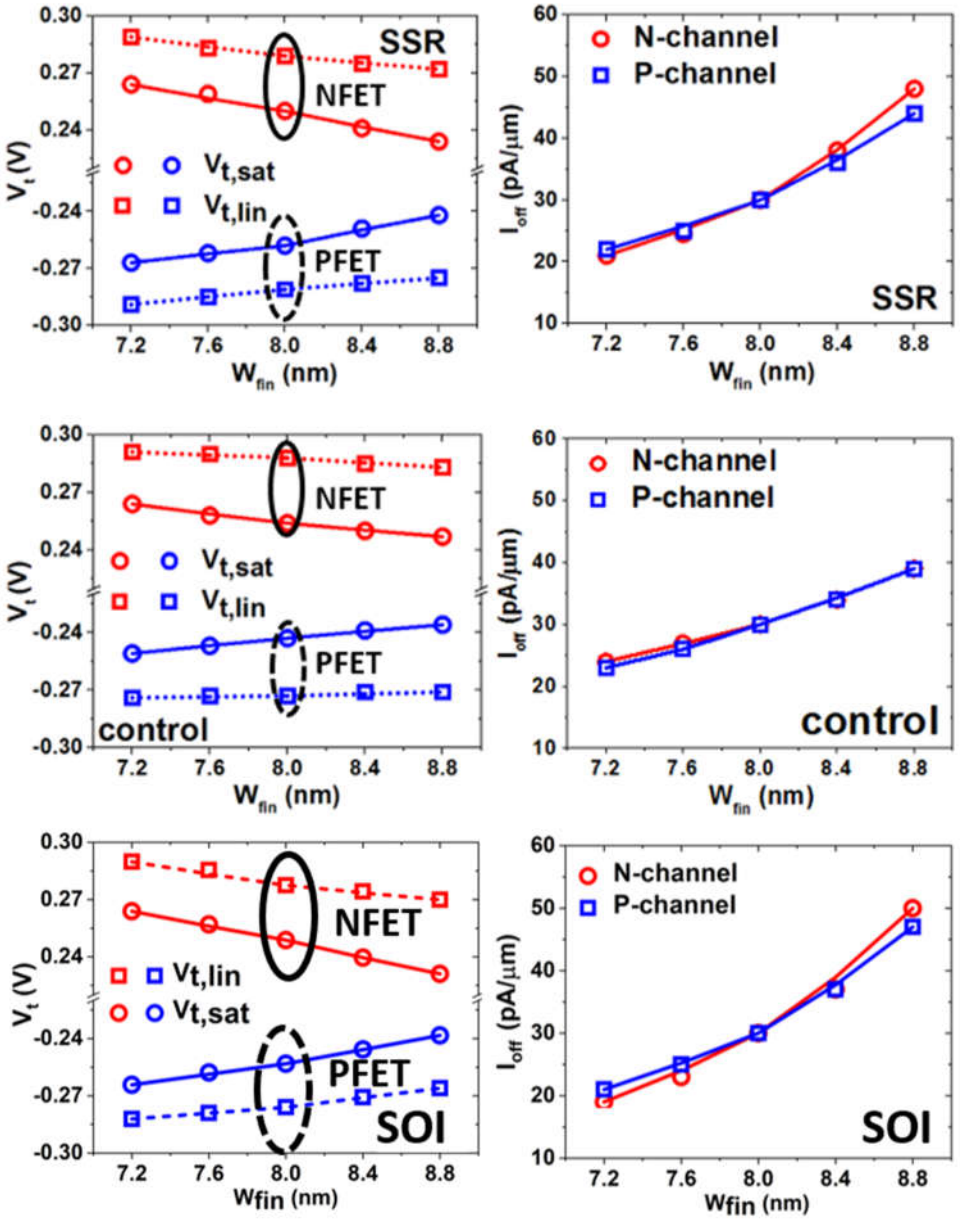

Fig. 4.12. Effects of fin-width (W_{fin}) variation on FinFET threshold voltage (V_{th}) and OFF-state leakage current (I_{off}).

In summary, SOI FinFET technology has slightly greater sensitivity of I_{off} to L_{gate}, W_{Si} variations because: (1) SOI FinFET has a slightly steeper subthreshold slope, and $\log(I_{off})$ is proportional to $\frac{V_{th}}{\text{SSwing}}$. (2) SOI FinFET technology relies solely on a narrow fin to suppress SCEs.

4.3.2 Impact of Random Variations

In this work, two random sources of variations are considered: random dopant fluctuations (RDF), and gate work function variation (WFV). Previous work has identified WFV as the dominant contributor to V_{th} variation for FinFET technology [35]. In this work, the gate material is assumed to be TiN with work function distributions taken from [30]. Depending on the average dopant concentration, variations in V_{th} and I_{off} due to RDF can become significant as the volume of the fin channel region shrinks [36]. Thanks to the employment of spacer lithography [37] to define nanometer-scale critical dimensions (L_{gate} and W_{fin}), gate line-edge-roughness (LER) is not expected to be a significant source of random variability in FinFET performance [38]. (In a self-aligned double patterning process, the critical dimension is defined by the thickness of a deposited film, which is locally very uniform.)

Random Dopant Fluctuations (RDF) Significant variation in threshold voltage can be caused by RDF for planar bulk-Si MOSFETs with L_{gate} less than 0.1 μm because of the tiny volume of the depletion region resulting in a relatively small amount of dopant atoms which determine V_{th} [13]. The standard deviation of V_{th} deviation is proportional to $\sqrt{\frac{N_a}{WL}}$, where N_a is the average dopant concentration in the depletion region, W is the channel width, and L is the channel length. In this work, RDF-induced variability in transistor performance is determined using the noise-like impedance field method [33, 34]. The results summarized in Table 4.3 show that both SSR and SOI FinFET technologies are barely impacted by RDF since they have relatively light dopant concentration within the (fully depleted) fin channel region so that their depletion charge negligibly affect V_{th}.

Work Function Variation (WFV) To maintain gate control and suppress SCE, the equivalent oxide thickness (EOT) of the gate dielectric layer(s) need to be scaled commensurately with transistor gate length. However, a high-permittivity (high-k) dielectric needs to be used in conjunction with SiO_2 to avoid excessive gate leakage due to direct tunneling through an ultra-thin dielectric for MOSFETs with EOT $<$ 1 nm. Fermi-level pinning at the interface of doped polycrystalline-silicon (poly-Si) and high-k material undesirably affects the effective work function of the poly-Si. In addition, remote soft optical phonon scattering degrades inversion-layer mobility significantly [39]. Therefore, a metal gate material must be used together with a high-k dielectric material. High-k/metal gate stacks have been used in mass production of CMOS since the introduction of Intel's 45 nm technology [40].

The work function (WF, in eV) is defined as the minimum energy required to remove an electron from the solid material, which equals the sum of bulk chemical potential (due to electron-electron correlation and exchange effects) and surface dipole potential. The bulk chemical potential is a fixed material property, whereas the surface dipole potential

depends on crystalline orientation. Due to the statistical nature of the metal layer deposition process, WF of a metal gate electrode suffers from local variations.

In this work, the random sources of variation (WFV and RDF) are assumed to be independent so that the intrinsic V_{th} variation can be calculated as:

$$\sigma V_{\mathrm{t,\ total}} = \sqrt{\sigma V_{\mathrm{t,\ RDF}}{}^{2} + \sigma V_{\mathrm{t,\ WFV}}{}^{2}} \tag{4.1}$$

Table 4.3 summarizes the impacts of random sources of variations on $\sigma V_{\mathrm{t,\ sat}}$, $\sigma I_{\mathrm{d,\ sat}}$, and σI_{off}. Note that WFV has a dominant effect. Thanks to a lightly doped fin channel region (1×10^{15} cm^{-3}), both SSR and SOI FinFET technologies can be seen to be barely affected by the RDF. Overall, both SSR and SOI FinFET technologies are shown to be relatively immune to random sources of variations compared with conventional heavily doped bulk-Si (control) FinFETs.

		Control FinFETs		SSR FinFETs		SOI FinFETs	
		n-channel	p-channel	n-channel	p-channel	n-channel	p-channel
$\sigma V_{\mathrm{t,\ sat}}$ (V)	RDF-induced	16	13.4	2.1	2.5	1.21	0.86
	WFV-induced	23.2	23.4	23.7	22.9	23.2	22.5
	Total	28.2	27.0	23.8	23.0	23.2	22.5
$\sigma I_{\mathrm{d,\ sat}}$ ($\mu A/\mu m$)	RDF-induced	10.9	13.7	4.6	7.8	2.64	2.56
	WFV-induced	13.2	12.8	13.6	12.4	14.38	12.49
	Total	17.1	17.3	14.4	14.7	14.62	12.75
σI_{off} (p$A/\mu m$)	RDF-induced	8.3	7.1	1.9	1.5	1.22	0.96
	WFV-induced	13.8	12.3	13.7	11.6	13.8	11.8
	Total	15.9	14.2	13.8	11.7	13.9	11.8

Table 4.3. Random sources of variations induced variability in $V_{\mathrm{t,\ sat}}$, $I_{\mathrm{d,\ sat}}$ and I_{off}

4.4 6T-SRAM Performance and Yield

Reduction in 6T-SRAM cell variability is crucial to lowering the power supply voltage (V_{DD}) and hence reducing power consumption. In this section, two 6T-SRAM performance metrics, the read static noise margin (SNM) and the writability current (I_w), are first introduced. The calibrated compact model [41] is used instead of the computationally expensive mixed-mode TCAD device simulations to accurately calculate these metrics. To find the minimum 6T-SRAM cell operating voltage ($V_{DD,min}$), the sensitivities of SNM and I_w to variations in device parameters X_I, $\partial_{\mathrm{SNM}}/\partial X_I$, and $\partial I_w/\partial X_I$, are calculated to determine the minimum variability that cause either read or write failure as a function of the cell operating voltage.

4.4.1 6T-SRAM Cell Performance

An SRAM array consists of many cells (each cell stores one bit of information) arranged in rows and columns. Figure 4.13 schematically illustrates the 6T-SRAM cell architecture, which comprises two cross-coupled inverters. Each inverter consists of one p-channel "pull-up" (**PU**) transistor with its source tied to V_{DD} and one n-channel "pull-down" (**PD**) transistor with its source tied to ground. Two n-channel "pass-gate" (**PG**) transistors are used to connect the left and right internal storage nodes **CH** and **CL** to the left and right bit lines **BL** and $\overline{\mathbf{BL}}$, respectively. SRAM cells in the same column share the same **BL** and $\overline{\mathbf{BL}}$ bit lines. For each SRAM cell, a single wordline **WL** is used to control the **PG** transistors.

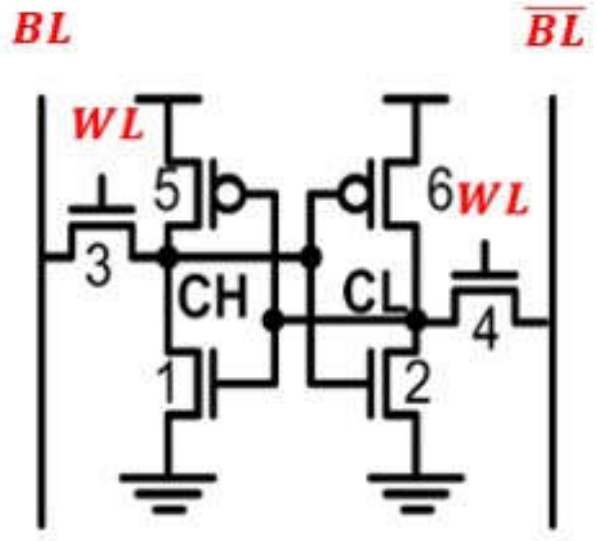

Figure 4.13. Circuit diagram of the 6T-SRAM cell design.

Read Static Noise Margin (SNM) The SNM is defined as the minimum amount of noise (change in voltage) required to disturb the cell, and quantitatively measures the robustness of an SRAM cell against a read disturb error. During a **Read** operation, both **BL** and $\overline{\textbf{BL}}$ are precharged to V_{DD}. Then **WL** is pulsed with a high voltage to turn on the **PG** transistors. Therefore, the bit lines are connected to the internal storage nodes, so that the bit line connected to the internal node storing a logic "0" (with a low voltage) is discharged through its corresponding **PG** and **PD** transistors, as shown in Figure 4.14. The SNM is extracted from the "butterfly plot" of the voltage transfer curves (V_{CH} *vs.* V_{CL}, and V_{CL} *vs.* V_{CH}) for the cross-coupled inverters during a **Read** operation, as illustrated in Figure 4.15.

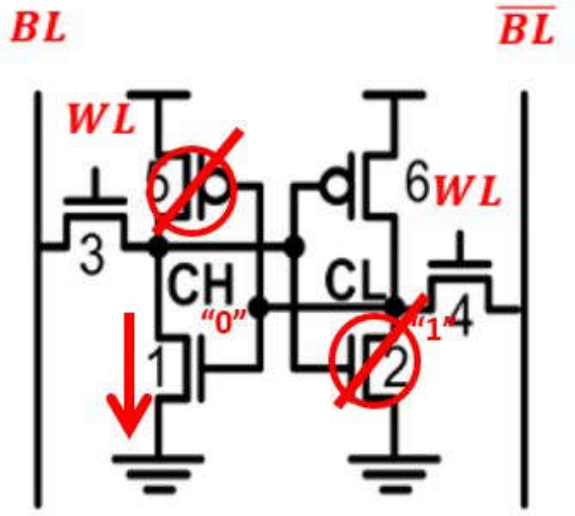

Fig. 4.14. Current flow in a 6T-SRAM cell during a Read operation. The **CH** node stores a logic "0" so that the **BL** is discharged through **PG** transistor 3 and **PD** transistor 1. If the voltage at CH rises above the tipping point of the opposite inverter so that **PG** transistor 2 turns on, it can flip to the "1" state erroneously [15].

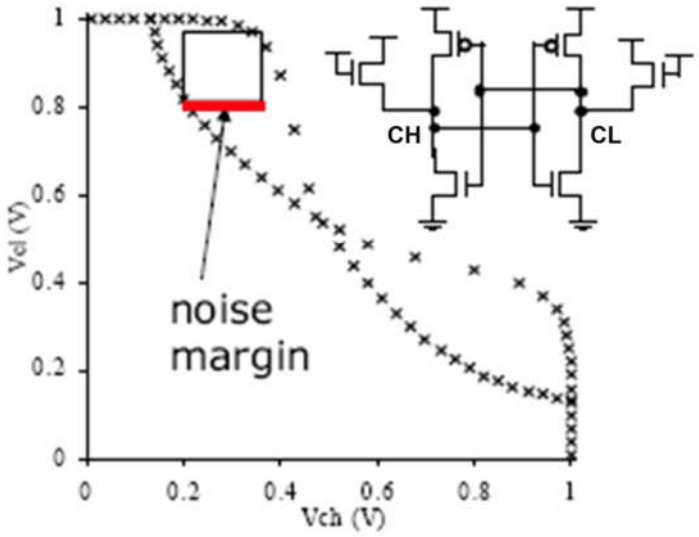

Fig. 4.15. The Read Static Noise Margin (SNM) corresponds to the length of the largest square that can be fitted within the smaller "lobe" of the butterfly plot [15].

To ensure a successful **Read** operation, it is desirable to have a larger cell **beta ratio**, defined as the ratio of **PD** transistor ON-state current to **PG** transistor ON-state current, such that the **PD** transistor has less ON-state resistance than the **PG** transistor. For conventional planar MOSFET technology, the cell **beta ratio** can be finely tuned by changing the drawn channel widths of the **PD** and **PG** transistors. However, for FinFET technology, the cell **beta ratio** can only be adjusted coarsely by adjusting the number of fins (connected in parallel between source and drain regions) in each device. Therefore, strong read stability is achieved by employing more fins for the **PD** devices than for the **PG** devices for FinFET technology.

Writability current (I_w) C. Wann of the IBM group proposed I_w [42] as a quantitative gauge to assess the immunity of the SRAM cell against write failure. During a **Write** operation, then **WL** is pulsed with a high voltage to turn on the **PG** transistors. To write information into internal storage nodes from **BL** and $\overline{\textbf{BL}}$ (carrying complementary logic values, *i.e.,* high and low voltages), the bit line at low voltage will discharge the internal storage node through the corresponding **PG** transistor, as illustrated in Figure 4.16. The **PG** transistor must be stronger than its corresponding **PU** transistor which tries to retain a high voltage on the internal storage node. The I_w is the minimum amount of current flowing out of this internal storage node during the discharging from V_{DD} toward ground potential, as shown in Figure 4.17. The cell **gamma ratio** is defined as the ratio of **PG** transistor ON-state current to **PU** transistor ON-state current. A larger **gamma ratio** is desirable for strong robustness against write failure. For planar MOSFET technology,

this can be achieved by adjusting the drawn channel widths of the **PG** and **PU** transistors. However, the **gamma ratio** can be only tuned coarsely by adjusting the number of fins (connected in parallel between source and drain regions) in each device.

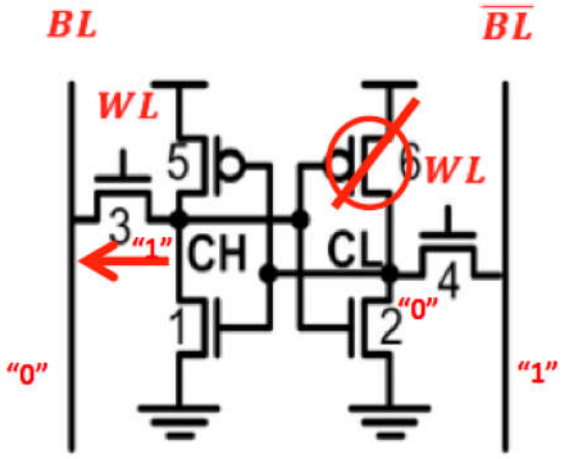

Fig. 4.16. Current flow in a 6T-SRAM cell during a Write operation. The **CH** node is storing a logic "1" and must be discharged through **PG** transistor 3; hence is resisted by **PU** transistor 5. A write failure occurs if the **PU** transistor is stronger than the **PG** transistor.

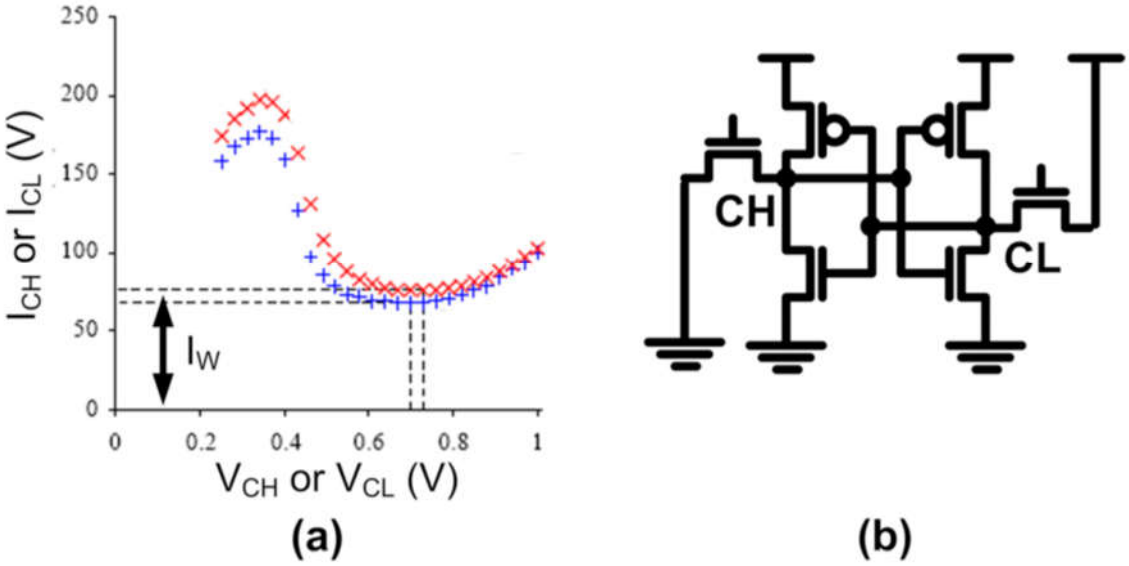

Fig. 4.17. The Writability current (I_w) of a 6T-SARM cell corresponds to the local minimum of either the "write-N" curves I_{CH} *vs.* V_{CH} and I_{CL} *vs.* V_{CL}. These curves are generated by sweeping V_{CH} or V_{CL} and measuring the nodal current at **CH** or **CL**, respectively, during a Write operation.

4.4.2 FinFET-based 6T-SRAM Cell Designs

For FinFET technology, the 6T-SRAM cell gamma ratio and beta ratio can only be tuned coarsely by adjusting the number of fins (connected in parallel between source and drain regions) in each device. It can be seen from Table 4.2 that pFETs have higher drive current as compared with nFETs because the hole mobility is higher on {110} fin sidewall surfaces [43], uniaxial stress is more effective for boosting hole mobility [44], and the performance benefits of embedded-SiGe S/D regions for pFETs [45]. Previous research showed that the performance improvements in pFETs (**PU** transistors) degrade 6T-SRAM write-ability, thus stronger **PG** devices are needed to counteract this effect [6]. In turn, better read stability can be ensured by using more fins for the **PD** transistors than for the **PG** transistors. In this section, the performance and yield of 1-1-1, 1-2-2, and 1-3-3 6T-SRAM cell designs are studied via calibrated compact model mentioned above, which are consistent with TSMC's 16-nm FinFET technology [46]. The 1-3-3 cell design comprises 1-fin PU, 3-fin PD, and 3-fin PG devices.

Compact Modeling of Read SNM & I_w Read stability and write-ability are gauged by the read SNM and I_w metrics derived from the butterfly plot and write N-curve generated using the calibrated compact model. Figure 4.18 shows the modeled butterfly curves for 1-3-3 FinFET 6T-SRAM cell design. The read SNM for the SSR FinFET technology is comparable to that for the control FinFET technology. An SOI FinFET technology has slightly greater read SNM due to stronger **PD** nFETs compared with an SSR FinFET technology. Read SNM is generated from the inverter VTCs obtained by applying Kirchhoff's Current Law (KCL), for example:

$$\begin{aligned} I_{D1}(V_{GS} = V_{CL}, V_{DS} = V_{CH}) = I_{D3}(V_{GS} = V_{WL} - V_{CH}, V_{DS} = V_{BL} - V_{CH}) \\ + I_{D5}(V_{GS} = V_{CL} - V_{DD}, V_{DS} = V_{CH} - V_{DD}) \end{aligned} \tag{4.2}$$

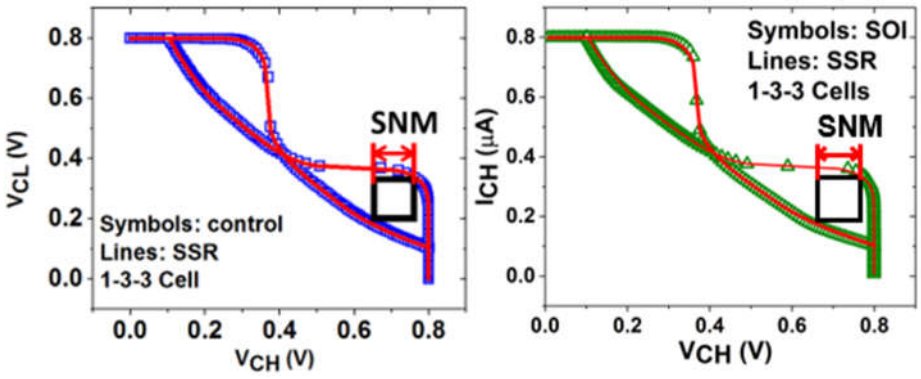

Fig. 4.18. Modeled butterfly curves for 1-3-3 6T-SRAM bit cells with bulk-Si FinFET technology (left) and SOI FinFET technology (right). Lines represent compact model simulated butterfly curves for SSR FinFET technology in both figures for comparison.

I_w is determined iteratively from calculating the I_{CH} *vs.* V_{CH} and I_{CL} *vs.* V_{CL} curves corresponding to write "0" and write "1" operations, respectively:

$$I_{CH} = I_{D3}(V_{GS} = V_{WL}, V_{DS} = V_{CH}) - I_{D5}(V_{GS} = V_{CL} - V_{DD}, V_{DS} = V_{CH} - V_{DD}) + I_{D1}(V_{GS} = V_{CL}, V_{DS} = V_{CH}) \quad (4.3)$$

Figure 4.19 shows the modeled write N-curves for 1-3-3 6T-SRAM bit cells. Because the SSR FinFET technology has a slightly larger gamma ratio (1.08) than the control FinFET technology (1.06), the SSR FinFETs provide for better write-ability. Table 4.4 summarizes the nominal read SNM and I_w for FinFET-based SRAM cells. Thanks to the stronger PD nFETs and slightly larger gamma ratio, the SOI FinFET technology has slightly greater read SNM and I_w as compared with the control FinFET technology and the SSR FinFET technology. Figure 4.20 compares the alpha ratio (*i.e.*, $\frac{1}{\text{gamma ratio}}$) of 7/8-nm SOI/SSR FinFET technologies studied in this work with that for previous technology nodes [46]. It is evident that the introduction of SiGe S/D regions in PFET increases alpha ratio (decreases gamma ratio) by boosting PFET $I_{d,sat}$, resulting in a decrease in write-ability.

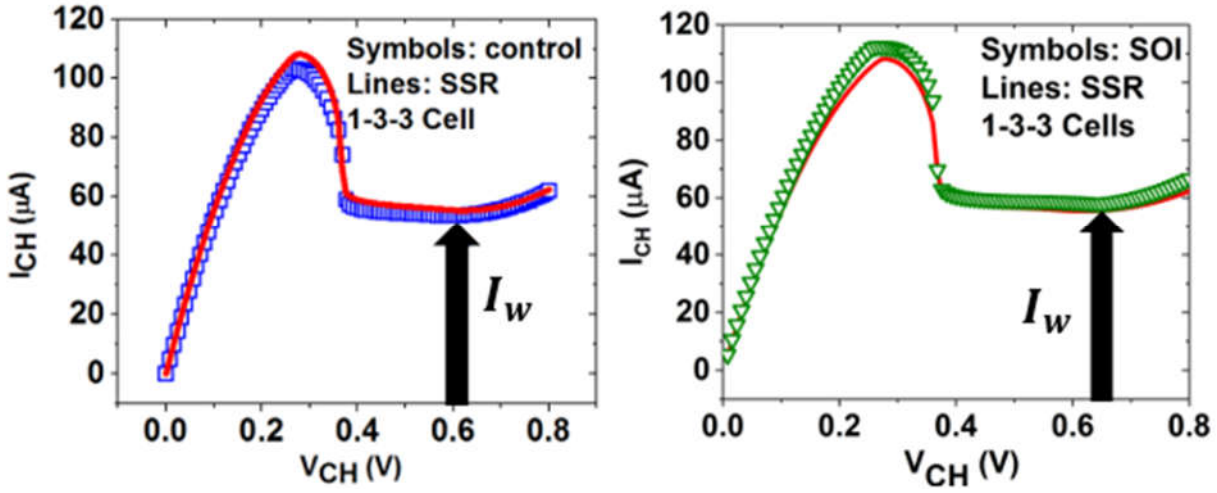

Fig. 4.19. Modeled write N-curves for 1-3-3 6T-SRAM bit cells with bulk-Si FinFET technology (left) and SOI FinFET technology (right). Lines represent compact model simulated write N-curves for SSR FinFET technology in both figures for comparison.

Cell Design	Control FinFETs		SSR FinFETs		SOI FinFETs	
	SNM (mV)	I_w (μA)	SNM (mV)	I_w (μA)	SNM (mV)	I_w (μA)
1-1-1	163	15	162	16	166	116
1-2-2	149	34	148	36	154	37
1-3-3	142	54	141	55	148	58

Table 4.4. Summary of FinFET-based 6T-SRAM cell performance metrics (at V_{DD} = 0.80 V)

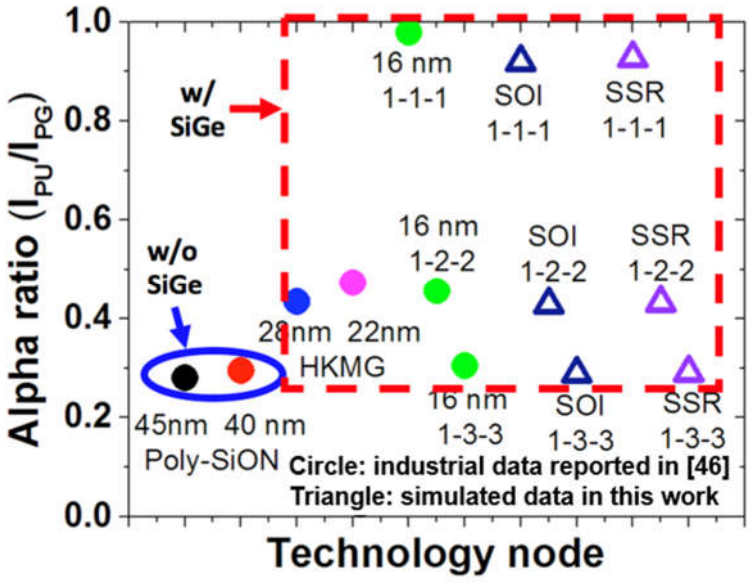

Fig. 4.20. Comparison of alpha ratio for simulated devices and reported technology (triangles: simulated 7/8-nm SOI and SSR FinFET technologies in this work; circles: data from [46]).

4.4.3 6T-SRAM Cell Yield Estimation

Systematic and random sources of variations induced transistor performance variability can result in read SNM < 0 V (read disturb) or $I_w < 0$ A (write failure) for a 6T-SRAM cell. Cell sigma is defined as the minimum total number of standard deviations from nominal values of L_{gate}, W_{fin}, and/or $V_{t,sat}$ for each of the 6 transistors in a 6T-SRAM cell,

that causes a read SNM < 0 V or $I_w < 0$ A. In this work, both process variations assuming 3σ deviation corresponding to $\pm 10\%$ variation in nominal L_{gate}, W_{fin} values and random sources of variations including RDF and WFV are considered to model cell sigma in a multi-dimensional variation space. In this variation space, each dimension corresponds to one device parameter, and the probability of occurrence decreases with increasing deviation from the nominal value. There exists a region corresponding to combinations of the transistor parameter variations which cause read disturb or write failure, referred to as the **surface of failure**. The read SNM cell sigma is defined as the shortest distance from the origin (nominal device parameter values) to this surface of failure.

Figure 4.21 shows both read SNM and I_w degrade with decreasing V_{DD}. Since SiGe S/D regions increase the alpha ratio significantly, write-margin degradation has become a particularly severe problem. Figure 4.22 directly compares cell sigmas of SSR FinFET vs. that of SOI FinFET SRAM bit cells, for 1-1-1, 1-2-2, and 1-3-3 cell designs.

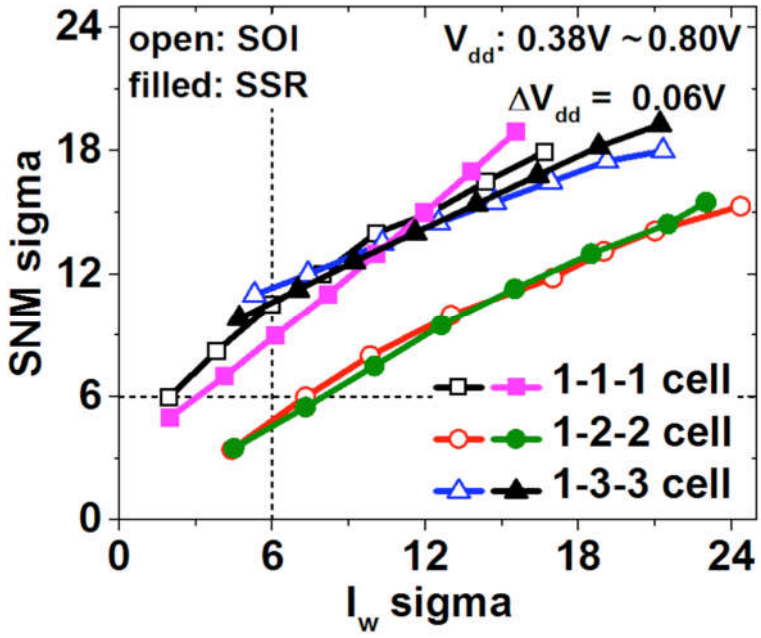

Fig. 4.21. Compact model predicted SNM, I_w *vs.* V_{DD}, V_{DD} is reduced from 0.80 V to 0.30 V in 50 mV step [2].

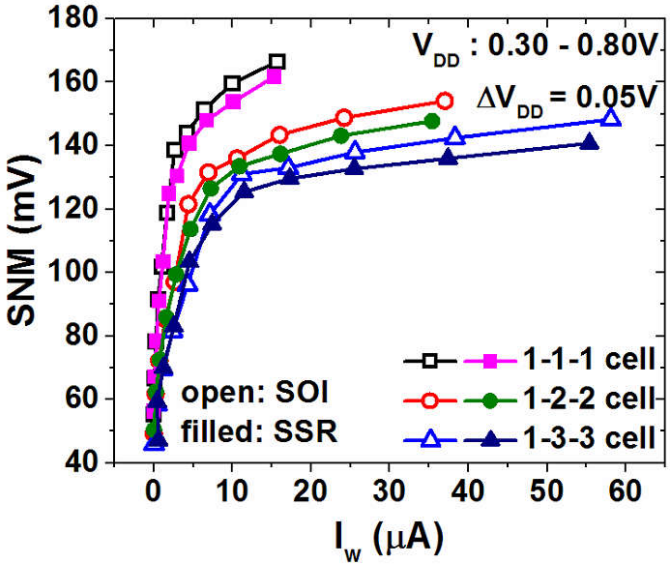

Fig. 4.22. Read SNM yield *vs.* write-ability current yield, for the 1-1-1, 1-2-2, and 1-3-3 6T-SRAM bit cells with SOI FinFETs or SSR FinFETs, V_{DD} is reduced from 0.80 V to 0.38 V in 60 mV step.

$V_{DD,min}$, the minimum 6T-SRAM cell operating voltage, is defined as the lowest V_{DD} that meets the six-sigma yield requirement for both Read and Write operations. Table 4.5 summarizes $V_{DD,min}$ for various 6T-SRAM cell designs. Lowest $V_{DD,min}$ is found to be 0.39 V and 0.40 V, respectively, for 1-3-3 bit cells implemented with SOI FinFETs and SSR FinFETs. 1-1-1, 1-2-2, and 1-3-3 bit cells implemented with SOI FinFETs and SSR FinFETs are projected to enable cell operating voltage $V_{DD,min}$ below 0.50 V. SOI FinFET technology only provides up to 20 mV (4.35%) reduction in $V_{DD,min}$ for 1-2-2 bit cells; SOI FinFET technology only provides up to 10 mV (2.5%) reduction in $V_{DD,min}$ for 1-3-3 bit cells as compared with SSR FinFETs. Both SSR and SOI FinFET technologies can be seen to provide significant improvement compared with heavily-doped bulk-Si (control) FinFET technology: 16.67% and 16.67% improvement for 1-1-1 bit cells; 18.5% and 15.0% for 1-2-2 bit cells; 15.2% and 13.0% improvement for 1-3-3 bit cells.

Cell Design	Control FinFETs $V_{DD,min}$ (V)	SSR FinFETs $V_{DD,min}$ (V)	SOI FinFETs $V_{DD,min}$ (V)
1-1-1	0.60	0.50	0.50
1-2-2	0.54	0.44	0.46
1-3-3	0.46	0.39	0.40

Table 4.5. Comparison of minimum 6T-SRAM cell operating voltage for various FinFET-based cell designs.

4.5 Summary

OI technology can facilitate the formation of a super-steep retrograde fin channel doping profile, which can overcome the scaling challenges faced by the conventional bulk-Si FinFET technology, achieving higher I_{on}/I_{off} ratios and reducing the sensitivity of device performance to systematic and random sources of variations. The SSR FinFET technology is benchmarked against SOI FinFET technology. Both SSR FinFET and SOI FinFET technologies can reduce random sources of variations induced $\sigma V_{\mathrm{t,sat}}$ below 25 mV, whereas SOI FinFET technology provides slightly stronger robustness against RDF. Thanks to the reduced transistor performance variations, both SSR and SOI FinFET technologies are projected to facilitate reductions in the minimum cell operating voltages (by as much as 100 mV compared with the control FinFET technology). However, due to the marginal improvement in mitigating transistor performance variations, SOI FinFET technology only provides for slightly smaller $V_{DD,min}$ for 1-2-2, and 1-3-3 bit cells, and same $V_{DD,min}$ for 1-1-1 bit cells. This study demonstrates that both SSR and SOI FinFET technologies can extend CMOS scaling beyond the 10-nm node, while SSR FinFET technology is prominent as a cheaper alternative to SOI FinFET technology for low-power 7/8-nm technology node.

4.6 References

[1] X. Zhang *et al.*, "Analysis of 7/8-nm bulk-Si FinFET technologies for 6T-SRAM scaling," *IEEE T Electron Dev,* vol. 63, no. 4, pp. 1502-1507, 2016.

[2] X. Zhang, D. Connelly, H. Takeuchi, M. Hytha, R. J. Mears, and T.-J. K. Liu, "Comparison of SOI versus bulk FinFET technologies for 6T-SRAM voltage scaling at the 7-/8-nm node," *IEEE T Electron Dev,* vol. 64, no. 1, pp. 329-332, 2017.

[3] X. Huang *et al.*, "Sub-50 nm P-channel FinFET," *IEEE T Electron Dev,* vol. 48, no. 5, pp. 880-886, 2001.

[4] C. Auth *et al.*, "A 10nm high performance and low-power CMOS technology featuring 3 rd generation FinFET transistors, Self-Aligned Quad Patterning, contact over active gate and cobalt local interconnects," in *Electron Devices Meeting (IEDM), 2017 IEEE International*, 2017: IEEE, pp. 29.1. 1-29.1. 4.

[5] C. Auth *et al.*, "A 22nm high performance and low-power CMOS technology featuring fully-depleted tri-gate transistors, self-aligned contacts and high density MIM capacitors," in *2012 Symposium on VLSI Technology (VLSIT)*, 2012: IEEE, pp. 131-132.

[6] S. Natarajan *et al.*, "A 14nm logic technology featuring 2nd-generation FinFET, air-gapped interconnects, self-aligned double patterning and a 0.0588 μm^2 SRAM cell size," in *2014 IEEE International Electron Devices Meeting*, 2014: IEEE, pp. 3.7. 1-3.7. 3.

[7] E.-Y. Jeong *et al.*, "High performance 14nm FinFET technology for low power mobile RF application," in *2017 Symposium on VLSI Technology*, 2017: IEEE, pp. T142-T143.

[8] S.-Y. Wu *et al.*, "A 16nm FinFET CMOS technology for mobile SoC and computing applications," in *2013 IEEE International Electron Devices Meeting*, 2013: IEEE, pp. 9.1. 1-9.1. 4.
[9] D. Hisamoto *et al.*, "FinFET-a self-aligned double-gate MOSFET scalable to 20 nm," *IEEE T Electron Dev,* vol. 47, no. 12, pp. 2320-2325, 2000.
[10] J. L. Hennessy and D. A. Patterson, *Computer architecture: a quantitative approach*. Elsevier, 2011.
[11] H. Bu, "FINFET technology a substrate perspective," in *IEEE 2011 International SOI Conference*, 2011: IEEE, pp. 1-27.
[12] M. Jurczak, N. Collaert, A. Veloso, T. Hoffmann, and S. Biesemans, "Review of FINFET technology," in *2009 IEEE International Soi Conference*, 2009: IEEE, pp. 1-4.
[13] A. Asenov, "Random dopant induced threshold voltage lowering and fluctuations in sub-0.1μm MOSFET's: A 3-D" atomistic" simulation study," *IEEE T Electron Dev,* vol. 45, no. 12, pp. 2505-2513, 1998.
[14] N. Xu *et al.*, "Extension of Planar Bulk n-Channel MOSFET Scaling With Oxygen Insertion Technology," *IEEE T Electron Dev,* vol. 61, no. 9, pp. 3345-3349, 2014, doi: 10.1109/TED.2014.2342496.
[15] A. E. Carlson, *Device and circuit techniques for reducing variation in nanoscale SRAM*. University of California at Berkeley, 2008.
[16] L. Wilson, "International technology roadmap for semiconductors (ITRS)," *Semiconductor Industry Association,* vol. 1, 2013.
[17] A. Hokazono, K. Ohuchi, K. Miyano, I. Mizushima, Y. Tsunashima, and Y. Toyoshima, "Source/drain engineering for sub-100 nm CMOS using selective epitaxial growth technique," in *International Electron Devices Meeting 2000. Technical Digest. IEDM (Cat. No. 00CH37138)*, 2000: IEEE, pp. 243-246.
[18] S. Sun *et al.*, "Enable abrupt junction and advanced salicide formation with dynamic surface annealing," *physica status solidi (c),* vol. 9, no. 12, pp. 2436-2439, 2012.
[19] M. V. Fischetti and S. E. Laux, "Band structure, deformation potentials, and carrier mobility in strained Si, Ge, and SiGe alloys," *Journal of Applied Physics,* vol. 80, no. 4, pp. 2234-2252, 1996.
[20] N. Xu *et al.*, "MOSFET performance and scalability enhancement by insertion of oxygen layers," in *2012 International Electron Devices Meeting*, 2012: IEEE, pp. 6.4. 1-6.4. 4.
[21] Synopsys, *Sentaurus Device User Guide, Version G*. Mountain View, CA, 2017.
[22] M. Ancona and H. Tiersten, "Macroscopic physics of the silicon inversion layer," *Physical Review B,* vol. 35, no. 15, p. 7959, 1987.
[23] K.-H. Kao, A. S. Verhulst, W. G. Vandenberghe, B. Soree, G. Groeseneken, and K. De Meyer, "Direct and indirect band-to-band tunneling in germanium-based TFETs," *IEEE T Electron Dev,* vol. 59, no. 2, pp. 292-301, 2011.
[24] V. Vidya, *Thin-body silicon FET devices and technology*, University of California at Berkeley, 2007.
[25] A. J. Bhavnagarwala, X. Tang, and J. D. Meindl, "The impact of intrinsic device fluctuations on CMOS SRAM cell stability," *IEEE journal of Solid-state circuits,* vol. 36, no. 4, pp. 658-665, 2001.

[26] X. Wang, A. R. Brown, B. Cheng, and A. Asenov, "Statistical variability and reliability in nanoscale FinFETs," in *2011 International Electron Devices Meeting*, 2011: IEEE, pp. 5.4. 1-5.4. 4.

[27] X. Wang, B. Cheng, A. R. Brown, C. Millar, and A. Asenov, "Statistical variability in 14-nm node SOI FinFETs and its impact on corresponding 6T-SRAM cell design," in *2012 Proceedings of the European Solid-State Device Research Conference (ESSDERC)*, 2012: IEEE, pp. 113-116.

[28] H. F. Dadgour, K. Endo, V. K. De, and K. Banerjee, "Grain-orientation induced work function variation in nanoscale metal-gate transistors—Part I: Modeling, analysis, and experimental validation," *IEEE T Electron Dev,* vol. 57, no. 10, pp. 2504-2514, 2010.

[29] H. F. Dadgour, K. Endo, V. K. De, and K. Banerjee, "Grain-orientation induced work function variation in nanoscale metal-gate transistors—Part II: Implications for process, device, and circuit design," *IEEE T Electron Dev,* vol. 57, no. 10, pp. 2515-2525, 2010.

[30] X. Zhang *et al.*, "Physical model of the impact of metal grain work function variability on emerging dual metal gate MOSFETs and its implication for SRAM reliability," in *2009 IEEE International Electron Devices Meeting (IEDM)*, 2009: IEEE, pp. 1-4.

[31] K. J. Kuhn, "Reducing variation in advanced logic technologies: Approaches to process and design for manufacturability of nanoscale CMOS," in *2007 IEEE International Electron Devices Meeting*, 2007: IEEE, pp. 471-474.

[32] S. Xiong and J. Bokor, "Sensitivity of double-gate and FinFET Devices to process variations," *IEEE T Electron Dev,* vol. 50, no. 11, pp. 2255-2261, 2003.

[33] A. Wettstein, O. Penzin, E. Lyumkis, and W. Fichtner, "Random dopant fluctuation modelling with the impedance field method," in *International Conference on Simulation of Semiconductor Processes and Devices, 2003. SISPAD 2003.*, 2003: IEEE, pp. 91-94.

[34] K. El Sayed, E. Lyumkis, and A. Wettstein, "Modeling statistical variability with the impedance field method," in *Proc. Int. Conf. Simulation Semiconductor Process. Devices (SISPAD)*, 2012, pp. 205-208.

[35] T. Matsukawa *et al.*, "Comprehensive analysis of variability sources of FinFET characteristics," in *2009 Symposium on VLSI Technology*, 2009: IEEE, pp. 118-119.

[36] A. Dixit *et al.*, "Impact of stochastic mismatch on measured SRAM performance of FinFETs with resist/spacer-defined fins: Role of line-edge-roughness," in *2006 International Electron Devices Meeting*, 2006: IEEE, pp. 1-4.

[37] C. Bencher, Y. Chen, H. Dai, W. Montgomery, and L. Huli, "22nm half-pitch patterning by CVD spacer self alignment double patterning (SADP)," in *Optical Microlithography XXI*, 2008, vol. 6924: International Society for Optics and Photonics, p. 69244E.

[38] N. Seoane *et al.*, "Comparison of fin-edge roughness and metal grain work function variability in InGaAs and Si FinFETs," *IEEE T Electron Dev,* vol. 63, no. 3, pp. 1209-1216, 2016.

[39] M. V. Fischetti, D. A. Neumayer, and E. A. Cartier, "Effective electron mobility in Si inversion layers in metal–oxide–semiconductor systems with a high-κ insulator: The role of remote phonon scattering," *Journal of Applied Physics,* vol. 90, no. 9, pp. 4587-4608, 2001.

[40] K. Mistry *et al.*, "A 45nm logic technology with high-k+ metal gate transistors, strained silicon, 9 Cu interconnect layers, 193nm dry patterning, and 100% Pb-free packaging," in *2007 IEEE International Electron Devices Meeting*, 2007: IEEE, pp. 247-250.

[41] X. Zhang, *Simulation-based study of super-steep retrograde doped bulk FinFET technology and 6T-SRAM yield*, University of California at Berkeley, 2016.

[42] C. Wann *et al.*, "SRAM cell design for stability methodology," in *IEEE VLSI-TSA International Symposium on VLSI Technology, 2005.(VLSI-TSA-Tech).* 2005: IEEE, pp. 21-22.

[43] H. Irie, K. Kita, K. Kyuno, and A. Toriumi, "In-plane mobility anisotropy and universality under uni-axial strains in nand p-MOS inversion layers on (100),[110], and (111) Si," in *IEDM Technical Digest. IEEE International Electron Devices Meeting, 2004.*, 2004: IEEE, pp. 225-228.

[44] L. Smith *et al.*, "Exploring the limits of stress-enhanced hole mobility," *IEEE Electron Device Letters,* vol. 26, no. 9, pp. 652-654, 2005.

[45] H. Chen, D. Chidambarrao, S. Panda, S.-H. Oh, H. K. Utomo, and W. A. Rausch, "In situ doped embedded sige extension and source/drain for enhanced PFET performance," ed: Google Patents, 2007.

[46] M. Yabuuchi *et al.*, "16 nm FinFET high-k/metal-gate 256-kbit 6T SRAM macros with wordline overdriven assist," in *2014 IEEE International Electron Devices Meeting*, 2014: IEEE, pp. 3.3. 1-3.3. 3.

Chapter 5

Schottky Barrier Height Modification via Oxygen Insertion Technology

In this chapter, the effects of oxygen-insertion (OI) technology and low-energy fluorine (F) implantation on the Schottky barrier height (Φ_{Bp}) of a Pt/Ti/p-type Si metal-semiconductor (M/S) contact are presented. Both oxygen-insertion (OI) layers and F demonstrated to reduce Φ_{Bp} due to Ti2p and Si2p binding energy shifts before forming gas anneal (FGA), and due to retarded Pt diffusion into Si (facilitating low-Φ_{Bp} Pt mono-silicide formation) during FGA. OI layers are found to be more effective than a F implant for reducing Φ_{Bp}, suggesting that BF_2 ion implantation should be avoided for the formation of p-type source/drain (S/D) regions in p-channel MOSFETs.

5.1 Introduction

To maintain the pace of transistor density scaling according to Moore's law, the contacted gate pitch (CGP) must be scaled by 0.7× for each new technology node. This is achieved by scaling down the channel length (L_{CH}), the gate-sidewall spacer length (L_{SP}), and the length of the S/D regions. This means that the M/S contact area must be scaled down, resulting in larger contact resistance (R_{CON}); also, the S/D junction depth must be scaled proportionately with L_{CH}, which also increases parasitic resistance ($R_{parasitic}$) that diminishes the performance gain brought by gate-length (L_{Gate}) scaling. $R_{parasitic}$ degrades transistor performance by lowering transconductance (g_m) and ON-state current ($I_{D,sat}$) as a result of reduced gate overdrive voltage ($V_{ov} = V_{gs} - V_{th}$), resulting in slower integrated circuit (IC) "chip" operating speed. Figure 5.1 schematically illustrates parasitic resistance components for an ultimately scaled CMOS transistor. The scaling ratio of each region represents a tradeoff. Shallower S/D regions result in larger S/D parasitic resistance (R_{SD}), contact resistance (R_{CON}), and silicide resistance (R_{SIL}). A shorter L_{CH} can help mitigate this issue, at the cost of higher parasitic gate resistance (R_{gate}) due to poorer replacement gate filling. As CMOS transistor dimensions are scaled down toward the atomic limit, R_{CON} is an increasingly significant performance limiter [1, 2]. To reduce R_{CON}, two aspects of the M/S contacts must be improved: S/D active dopant concentration (N_{SD})

and Schottky barrier height (Φ_B).

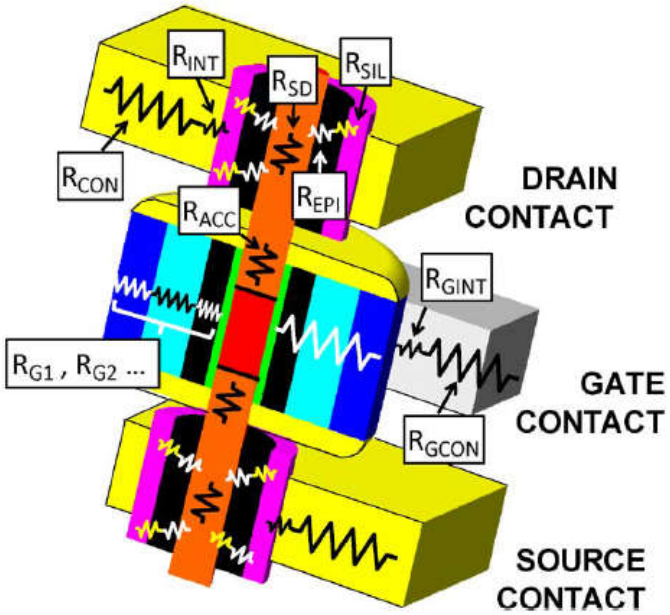

Fig. 5.1. Parasitic resistances components for an ultimately scaled MOS transistor (adapted from [1]).

Research efforts to reduce R_{CON} have followed different approaches : (1) increase the S/D active dopant concentration through solid phase epitaxial (SPE) recrystallization of implanted semiconductor regions [3-5]; (2) maximize the real contact area by removing highly-resistive residues using an optimized reactive-ion etching (RIE) cleaning process [6]; (3) reduce SBH (Φ_B) by inserting an ultra-thin interfacial layer (*i.e.*, forming an MIS structure) to de-pin the Fermi level [7-10] and/or by tuning the metal work-function [11, 12].

To reduce R_{CON} and thereby provide for improved CMOS transistor performance, the preferred choice of metal for S/D contact formation has changed over the past few decades. In the earliest CMOS technology, aluminum (Al) was commonly used to form Al/Si contacts [13]. It was replaced by $TiSi_2$ due to the Al spiking problem [14]. Lower resistivity $CoSi_2$ and Ni-Pt monosilicide were subsequently introduced to replace $TiSi_2$ [15]. With the advent of FinFET structures at the 22 nm technology node, the semiconductor industry returned to Ti-based contacts [16]. Ti has a low work-function value of 4.33 eV [17], resulting in a larger SBH ($\Phi_{Bp} \cong 0.70\,eV$) for p-type contacts and a smaller SBH ($\Phi_{Bn} \cong 0.42\,eV$) for n-type contacts [18]. In this work, the effects of OI layers and a low-energy F implant on the SBH of a Pt/Ti/p-type Si system (Φ_{Bp}) are investigated experimentally. The effects of OI layers and F on chemical-bond energy, metal/semiconductor intermixing, and metal-silicide formation during FGA are discussed.

5.2 Schottky Barrier Height Extraction

5.2.1 Schottky Diode Fabrication

To extract Φ_{Bp}, back-to-back Schottky diodes were fabricated on top of (100) p-type control and OI silicon wafer substrates. Some wafers received a 2.5 keV, 1×10^{15} cm^{-2} F implant with 7-degree tilt/0-degree rotation to study the effects of F. All wafers were diced into 2×2 cm^2 pieces. The chips were first cleaned in a sulfuric peroxide mixture (H_2SO_4: H_2O_2) bath at 120 °C for 10 minutes followed by a dip in dilute hydrofluoric acid (DHF) solution (10:1 H_2O:49% HF) until the surface became hydrophobic. Afterwards, the samples were immediately rinsed with de-ionized (DI) water. To recrystallize the amorphized Si lattice, F-implanted samples received a 1050 °C spike anneal in inert N_2 ambient following the chemical cleaning.

The OI samples have four partial monolayers of oxygen inserted into the crystalline Si substrate as shown in Figure 5.2. The 1.6 nm-thick Si cap layer serves to protect the OI layers from the strong O gettering effects of Ti [19].

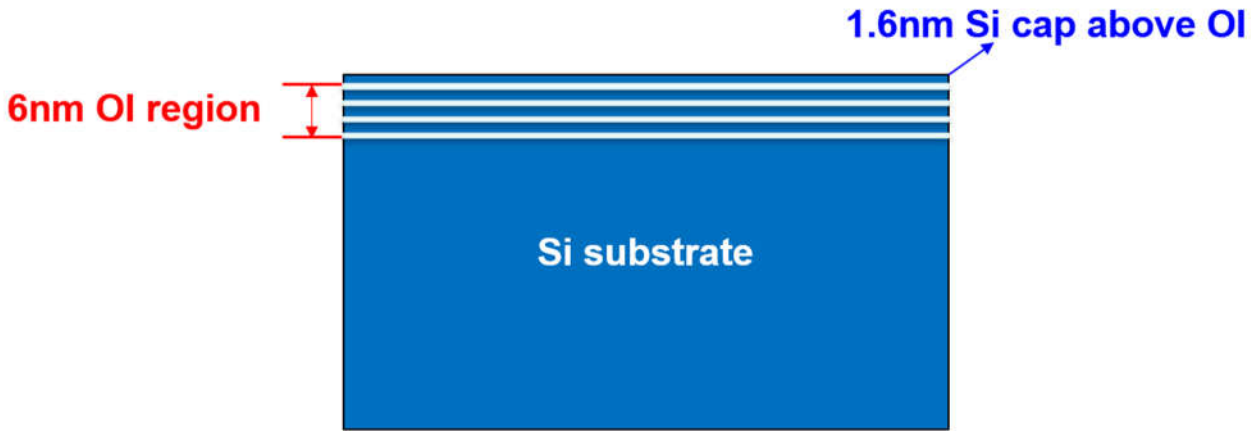

Fig. 5.2. Schematic cross-sectional view of the 6nm-thick OI region in the OI silicon wafer samples used in this work.

Figure 5.3 illustrates the fabrication process used to form back-to-back Schottky diode test structures on the sample (chip) surface, with r_1 ranging from 100 μm to 160 μm and r_2 ranging from 120 μm to 200 μm, respectively. First, a low-temperature (350°C) PECVD 230 nm-thick SiO_2 layer was deposited onto the cleaned substrate surfaces. This oxide layer subsequently serves to protect the Si surface from being etched by the TMAH-containing photoresist developer solution. Next, a 400 nm-thick lift-off resist (LOR-3A) layer and a 2 μm-thick g-line resist layer were spin-coated onto the sample surface in a Headway^{TM} photoresist spinner. Afterwards the bi-layer photoresist stack was patterned

by photolithography. Then the metal/Si contact regions were exposed by removing the LTO layer in DHF (10:1 H_2O:49% HF) solution for 7 minutes. Next, metal films (3 nm Ti followed by 10 nm Pt) were deposited by e-beam evaporation. The 3 nm Ti layer has superior interfacial adhesion while the Pt layer has ultrahigh conductivity and serves as a probe layer.

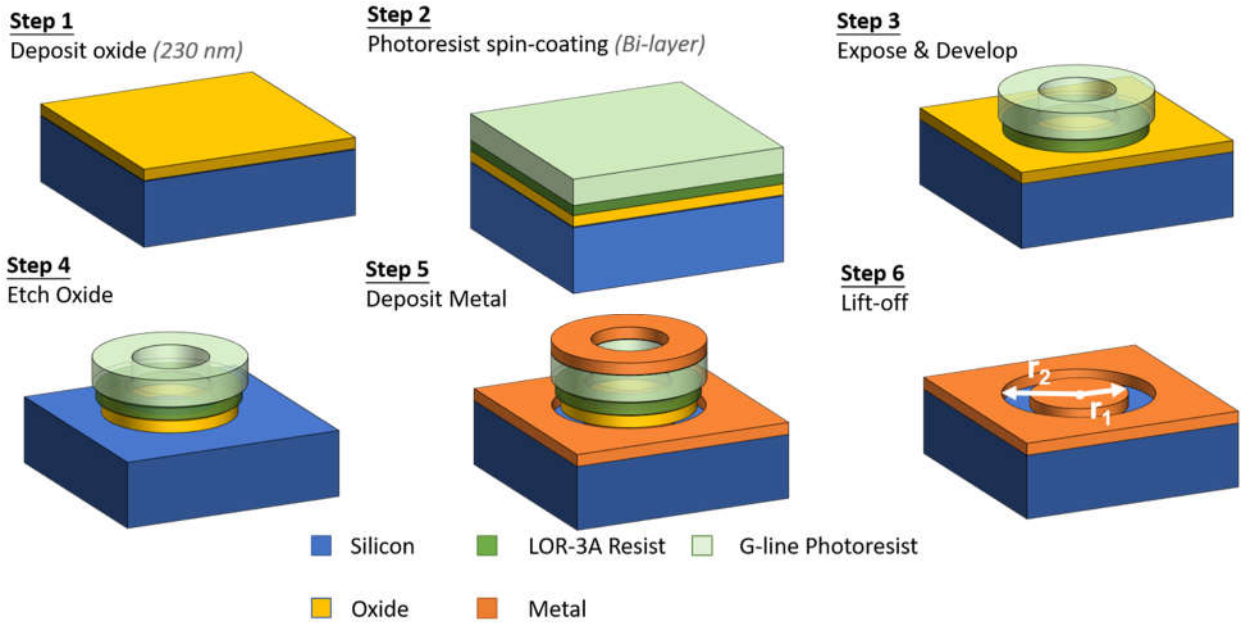

Fig. 5.3. Fabrication process flow for Schottky diode test structures.

5.2.2 Electrical Measurements

Based on thermionic emission theory [20], the reverse saturation current density (J_0) for a Schottky diode is given by:

$$J_0 = A^{**}T^2 exp(-q\Phi_B/k_BT) \tag{5.1}$$

where A^{**} is Richardson's constant, Φ_B is Schottky barrier height (SBH), and T is the absolute temperature in Kelvin.

It can be seen from Eq (5.1) that J_0 is exponentially dependent on Φ_B. This correlation allows Φ_B to be extracted from measurements of the temperature-dependent electrical current-voltage (IV) characteristics of Schottky diodes:

$$\Phi_B = \frac{k_BT}{q}[\ln(A^{**}) - \ln(\frac{J_0}{T^2})] \tag{5.2}$$

Figure 5.4 illustrates the experimental setup used for Schottky diode IV measurements in this work.

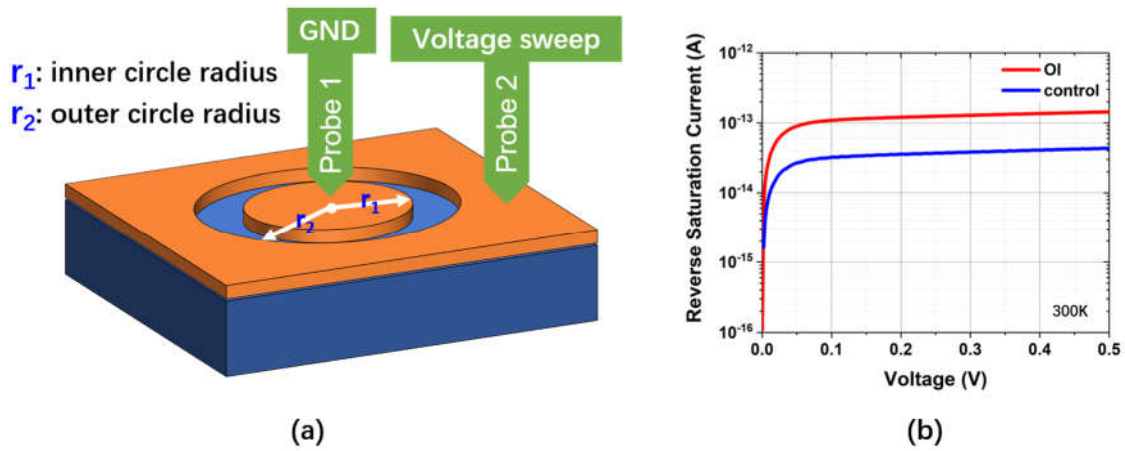

Fig. 5.4. Schottky diode IV measurement: (a) experimental setup, the applied voltage on probe 2 was swept from 0 V to 0.5 V with probe 1 grounded. (b) example IV curves for back-to-back Schottky diodes ($r_1 = 120\ \mu m$ and $r_2 = 140\ \mu m$), measured at 300 K.

Figure 5.5 shows Richardson plots ($\ln(\frac{I_r}{T^2})$ *vs.* $\frac{1000}{T}$) for control and OI samples within the temperature range of 220 K (-53.15 ℃) to 310 K (-36.85 ℃). The experimental data were fitted to straight lines using the least squares method, then the slope values were used to extract Φ_{Bp}. The extracted Φ_{Bp} value for the control sample (691 meV) is consistent with previously reported values for Ti/p-type Si contacts ($\Phi_{Bp} \cong 700$ meV). This suggests that the 3 nm-thick Ti layer continuously covers the semiconductor surface in the contact region. A comparison between the extracted Φ_{Bp} values for control and OI samples indicates that the OI layers result in lower Φ_{Bp} by 27 meV. Because of this, the Schottky diode saturation current (I_r) is higher for the OI sample. These results indicate that by inserting partial oxygen layers near to the Si substrate surface, R_{CON} of a p-type Ti/Si contact can be lowered.

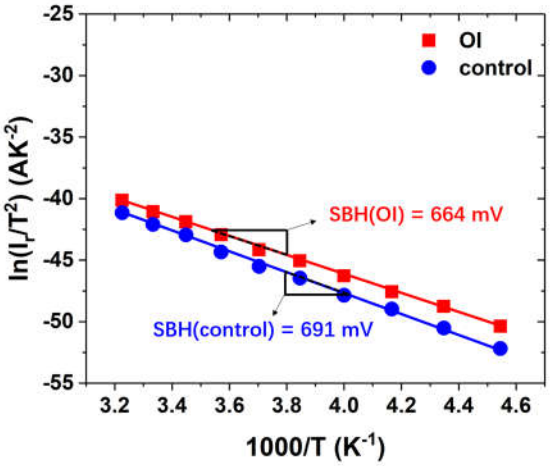

Fig. 5.5. Richardson plots for OI and control samples.

Thermal annealing in forming gas (a gaseous mixture of N_2 and H_2) is commonly used to passivate interface states for S/D contact formation to help reduce parasitic resistances [21, 22]. In this work, samples were subjected to a 5-minute, 300 ℃ anneal in 10% H_2, 90% N_2 forming gas in a conventional rapid thermal processing (RTP) system (AccuThermo^{TM} 610). Figures 5.6 and 5.7 show Richardson plots for un-implanted and F-implanted samples after FGA treatment. It can be seen that OI samples have lower Φ_{Bp} than control samples regardless of whether they had undergone F implant or FGA. Table 5.1 provides a summary of Φ_{Bp} values for all samples. A comparison of the values before *vs.* after FGA indicates that both OI layers and F can enhance FGA-induced Φ_{Bp} reduction. However, FGA can reduce Φ_{Bp} by more than 50% (664 meV → 330 meV) with the presence of OI layers, whereas FGA only reduces Φ_{Bp} by 3.7% (563 meV → 542 meV) with the presence of F. The OI FGA sample has the lowest Φ_{Bp} (330 meV) among all the samples, suggesting that OI technology is more beneficial than F implantation for reducing p-type contact resistance.

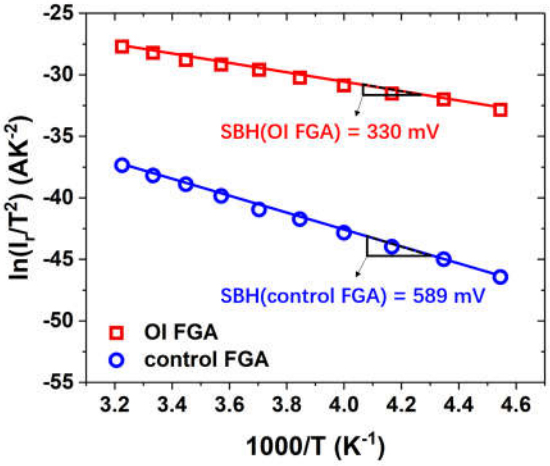

Fig. 5.6. Richardson plots for OI and control samples after FGA.

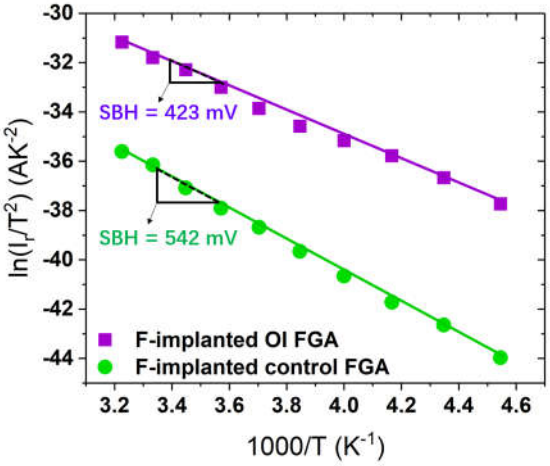

Fig. 5.7. Richardson plots for F-implanted OI and control samples after FGA.

Sample	Φ_{Bp} (meV)	
	Before FGA	After FGA
Un-implanted control	691	589
Un-implanted OI	664	330
F-implanted control	563	542
F-implanted OI	482	423

Table 5.1. Comparison of extracted Φ_{Bp} values for samples before and after FGA.

5.3 Oxygen Gettering Mechanism

To gain insight into the mechanisms for lower Φ_{Bp} in the presence of OI layers and F, secondary ion mass spectrometry (SIMS) and X-ray photoelectron spectroscopy (XPS) analyses were performed to elucidate the effects of oxygen and fluorine on Ti and Pt diffusion into Si, as well as on Pt silicidation.

Ti is a strong oxygen getter, and is typically used to remove O contaminants in standard CMOS fabrication processes [19]. At room temperature, Ti reacts with SiO_2 to form TiO_2: Ti + SiO_2 → TiO_2 + Si. The gettering process is accelerated at elevated temperatures. Figure 5.8 plots the O concentration profiles in OI samples before and after FGA. An O peak was observed in the top 3 nm-thick Ti layer even for the unannealed OI sample. This is likely due to the gettering of background O contamination during the e-beam evaporation process. A comparison between unannealed and annealed OI samples reveals that the OI layers lost some O atoms during FGA due to O gettering by Ti.

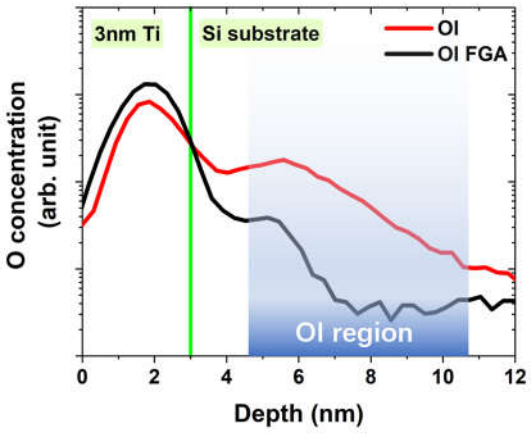

Fig. 5.8. Oxygen concentration profiles in OI samples before and after FGA, from SIMS analyses. The green line delineates the location of the Ti/Si interface. Figure 5.9 compares the O concentration profiles in control and OI samples before FGA. The O peak

in the Ti layer is similar for the control and OI samples.

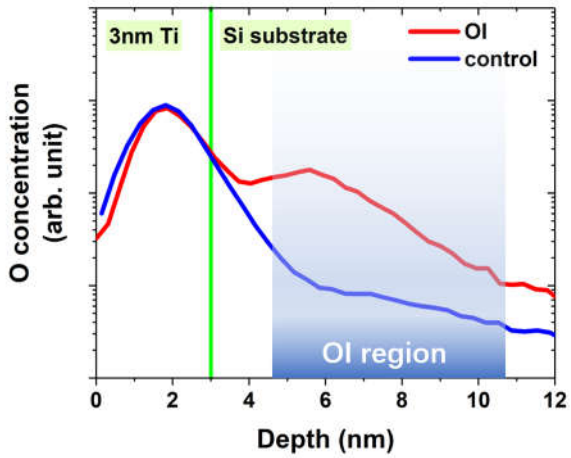

Fig. 5.9. Oxygen concentration profiles in OI and control samples before FGA, from SIMS analyses. The green delineates the location of the Ti/Si interface.

Figure 5.10 compares the O concentration profiles in control and OI samples after FGA. Due to FGA-enhanced O gettering, the O level in the Si substrate is much lower for the control sample than for the OI sample. This suggests that that Si-O bonds in the OI layers can help to retain O atoms, which can modify Φ_{Bp}.

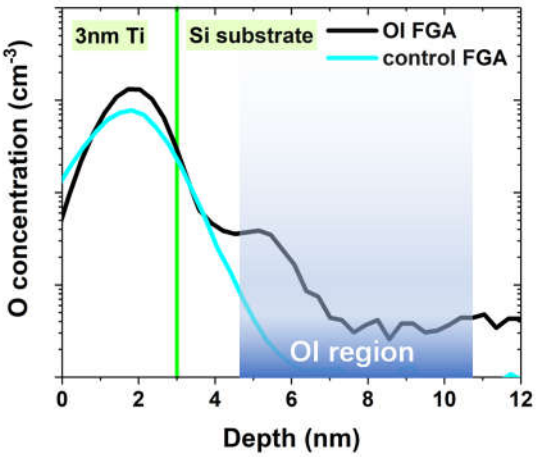

Fig. 5.9. Oxygen concentration profiles in OI and control samples after FGA, from SIMS analyses. The green line shows the location of the Ti/Si interface.

Previous studies identified the important role of interfacial chemistry to determine SBH [23-26]. In this work, XPS analyses were performed for the Pt/Ti/p-type Si system to study the interfacial chemical state and composition.

Table 5.2 lists the Ti $2p_{3/2}$ binding energies of common chemical states [27-29]. Higher values of Ti $2p_{3/2}$ binding energy are associated with Ti-O and Ti-N bonds. Therefore, it can be expected that O gettering of Ti (Ti-O bonding) results in a shift in the Ti $2p_{3/2}$

core energy level. Figures 5.11 and 5.12 plot XPS spectrum data of Ti 2p peaks at the Ti/Si interface for control and OI samples before FGA. Because of O gettering, the Ti $2p_{3/2}$ peak energy levels are higher than the ideal reference value (454.1 eV for pure Ti metal) [27] by 0.4 eV and 0.9 eV for control and OI samples, respectively. These values of Ti $2p_{3/2}$ core energy level are much smaller than that for TiO_2, indicating TiO_x ($x < 2$) formation at the Ti/Si interface. It can be expected that OI layers facilitate TiO_x formation since they are a source of additional O atoms, so that OI technology helps to change the Ti chemical state at the Ti/Si interface.

Chemical State	Binding Energy Ti $2p_{3/2}$ (eV)
Pure Ti metal	454.1
TiN	454.9
TiO_2	458.5

Table 5.2. Ti $2p_{3/2}$ binding energy values for common chemical states of Ti

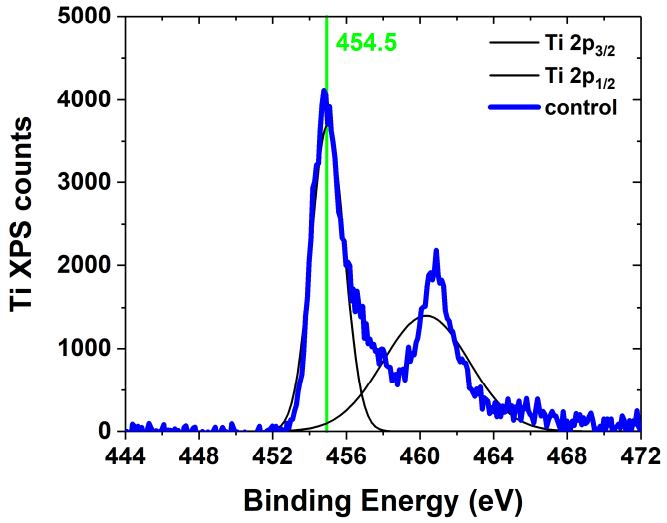

Fig. 5.11. XPS spectrum data for Ti 2p peaks (Ti $2p_{3/2}$ and Ti $2p_{1/2}$) for control sample before FGA.

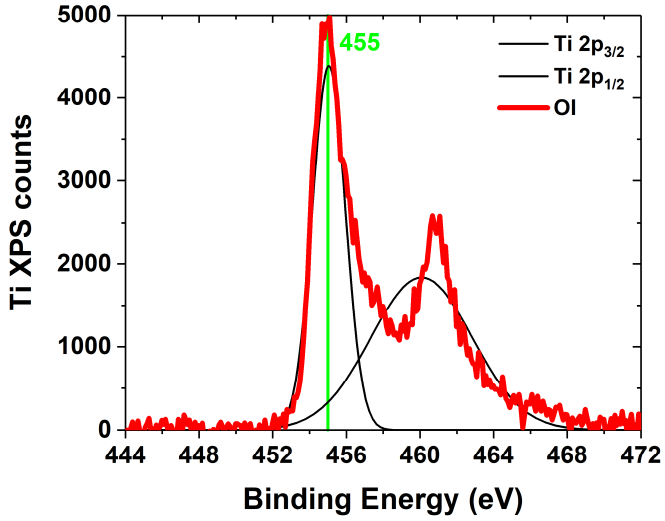

Fig. 5.12. XPS spectrum data of Ti 2p peaks (Ti $2p_{3/2}$ and Ti $2p_{1/2}$) for OI sample before FGA.

Previous work also discovered a correlation between the Ti chemical state and Ti/Si SBH [18]. It was hypothesized that an increase in Ti $2p_{3/2}$ core energy level corresponds to the work function of Ti moving closer to the Si valence band edge (E_v), which is beneficial for reducing Φ_{Bp}. The reduction in Φ_{Bp} by 27 meV (cf. Fig. 5.5) is consistent with the observed shift in the Ti $2p_{3/2}$ energy level due to Ti-O bond formation.

5.4 Pt Diffusion and Silicidation

Electrical measurements showed that FGA treatment can reduce Φ_{Bp} significantly, increasing back-to-back Schottky diode current flow for the Pt/Ti/p-type metal-silicon contact. The FGA thermal budget (5-minutes, 300 ℃) is insufficient to cause Ti silicidation, for which the lowest reported reaction temperature is 400 ℃ [30]. At room temperature, the 3 nm-thick Ti layer can serve as an effective barrier to Pt diffusion into the Si substrate to prevent Pt silicidation. During FGA, Pt diffusion across the thin Ti interlayer into the Si substrate can occur via a grain boundary enhanced diffusion mechanism [31], causing Pt silicide formation [32]. To understand the impacts of OI layers and F on Pt diffusion and Pt silicide formation, SIMS and XPS analyses were performed on samples that underwent FGA treatment.

5.4.1 Pt Diffusion into Si

Figure 5.13 plots the Pt concentration profiles for control and OI samples after FGA.

It can be seen that FGA causes a significant amount of Pt diffusion into the Si substrate for both samples. Previous work on Pt diffusion showed that O contamination (in the form of a 2 nm-thick discontinuous SiO_2 layer) at the Pt/Si interface can reduce Pt diffusion into the Si [32]. It was hypothesized that the thin oxide layer serves as a Pt diffusion barrier because the dominant Pt diffusion mechanism at low temperatures (200 – 325 ℃) is grain boundary diffusion. The difference in Pt concentration profiles for the control sample *vs.* the OI sample shows that OI layers can retard Pt diffusion into the Si substrate, *i.e.*, O atoms inserted interstitially into the crystalline Si lattice also can form a Pt diffusion barrier.

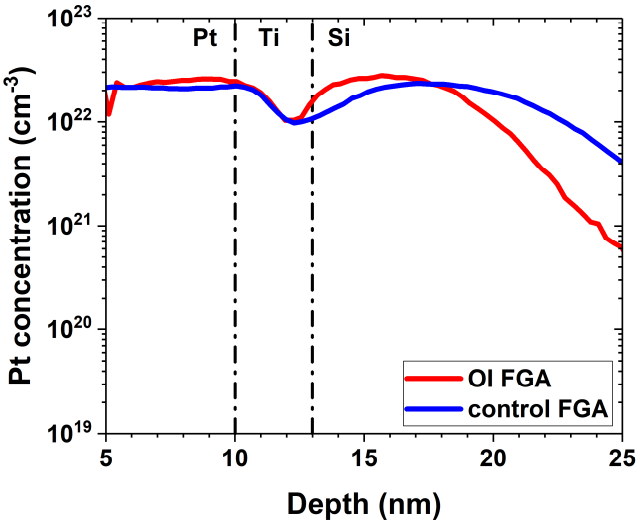

Fig. 5.13. Pt concentration profiles (log scale) from SIMS analyses, for control and OI samples after FGA. Notice that for the OI sample, Pt diffusion into Si is reduced as compared with the control sample.

Previous studies showed that the presence of F helps to reduce Pt diffusion into Si because of the "fluorine buffer" effect during high temperature (400 – 850 ℃) anneals [33, 34]. It is therefore expected that the presence of F reduces Pt diffusion into Si at lower temperatures. A comparison of Pt concentration profiles for un-implanted *vs.* F-implanted samples is shown in Figure 5.14. The F implant was found to reduce Pt diffusion into Si during FGA for both the control and the OI samples. However, the F-induced reduction in Pt diffusion is enhanced significantly in the presence of OI layers (F-implanted OI FGA sample). To elucidate why OI layers enhance the "fluorine buffer" effect, F concentration profiles before and after thermal anneals (the 1050 °C spike anneal prior to Schottky diode fabrication, and the 300°C FGA) are plotted in Figure 5.15. Due to high F diffusivity in silicon, the F concentration profile for the control FGA sample shows a typical "double-peak" consistent with previous studies of low-energy F implants [35]. However, this "double-peak" is not seen in the OI sample because F diffusion into the Si substrate is blocked by the OI layers [36, 37]. Previous work found that the F concentration must

exceed a certain level in order for F to effectively reduce Pt diffusion into Si [34]. Since the OI layers result in higher F concentration near the silicon surface, OI technology enhances the "fluorine-buffer" effects. This explains why the F-implanted OI FGA sample has the shallowest Pt depth profile among all F-implanted samples after FGA (*cf.* Figure 5.14).

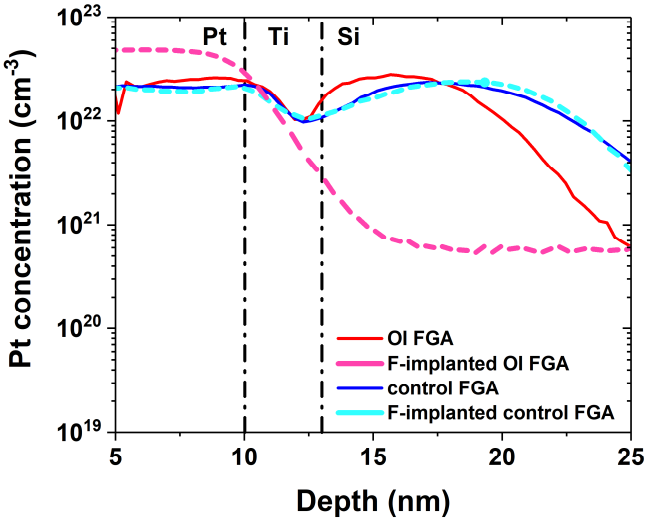

Fig. 5.14. Pt concentration profiles (log scale) from SIMS analyses for unimplanted (solid lines) and F-implanted (dashed lines) samples after FGA. Pt diffusion into the Si substrate is dramatically reduced in the presence of OI layers.

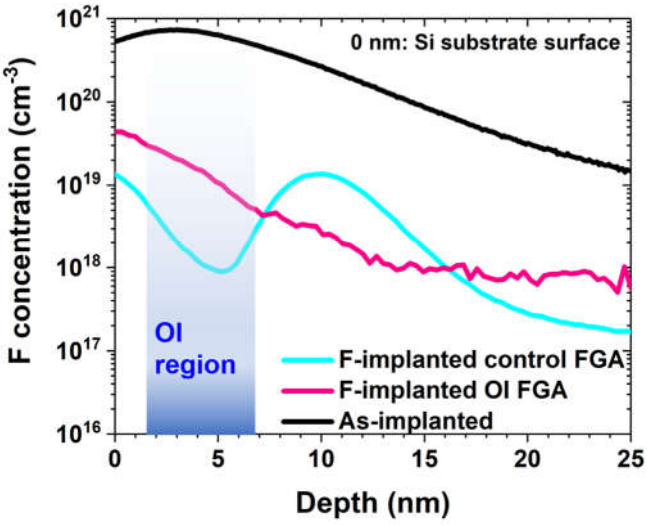

Fig. 5.15. F concentration profiles (log scale) from SIMS analyses before and after annealing (1050℃ recrystallization spike anneal and 5-minute 300℃ FGA).

5.4.2 Pt Silicide Phase

XPS analyses were performed to study the effects of OI layers and F on Pt silicidation during FGA. Table 5.3 provides a summary of common Pt $4f_{7/2}$ chemical states. Upon bonding with O (oxidation) and Si (silicidation), the Pt $4f_{7/2}$ core energy level shifts to larger values. Figures 5.16 and 5.17 plot XPS spectrum data for Pt 4f peaks at the original Si substrate surface for un-implanted samples after FGA. Based on the extracted Pt $4f_{7/2}$ binding energy shits (+1.11 eV for control FGA sample, +1.25 eV for OI FGA sample), it can be deduced that FGA caused Pt_2Si formation in the un-implanted control sample while it caused PtSi formation in the un-implanted OI sample. Previous research reported a PtSi/Si Φ_{Bp} of 320 meV [38], consistent with the extracted Φ_{Bp} of 330 meV for the OI FGA sample.

Chemical State	Binding Energy Pt $4f_{7/2}$ (eV)
Pure Pt metal	71.0
Pt_2Si	72.11
PtSi	72.25
PtO	72.4
PtO_2	74.9

Table. 5.3. Pt $4f_{7/2}$ binding energy values for common chemical states of Pt

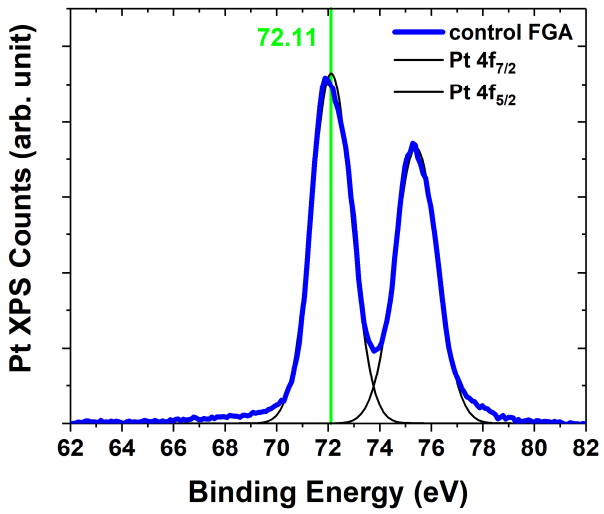

Fig. 5.16. XPS spectrum data of Pt 4f peaks for control FGA sample.

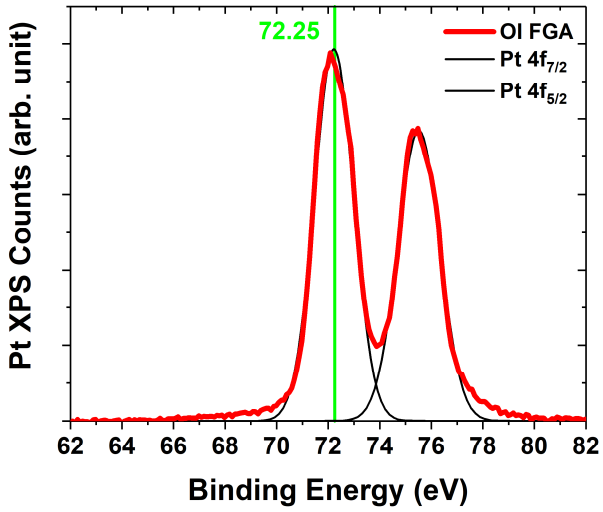

Fig. 5.17. XPS spectrum data of Pt 4f peaks for OI FGA sample.

Figures 5.18 and 5.19 plot XPS spectrum data of Pt 4f peaks at the original Si substrate surface for F-implanted samples after FGA. For the control sample, F increases the Pt $4f_{7/2}$ core energy level by 20 meV, from +1.11 eV to +1.13 eV. In contrast, F reduces the Pt $4f_{7/2}$ core energy level from +1.25 eV to +1.18 eV for the OI sample. Based on Table 5.3, these results suggest Pt_xSi ($1 < x < 2$) formation. It was shown in previous studies that a larger shift of the Pt $4f_{7/2}$ core level indicates more low-Φ_{Bp} Pt monosilicide formation, whereas a smaller shift indicates more high-Φ_{Bp} Pt-rich silicide formation [38]. Consistent correlations between Pt $4f_{7/2}$ core level shifts and Φ_{Bp} were observed in this work. For the control sample after FGA, F reduces Φ_{Bp} by 47 meV (589 meV → 542 meV); in contrast, for the OI sample after FGA, F increases Φ_{Bp} from 330 meV to 423 meV. These results show that Φ_{Bp} is very sensitive to the Pt silicide phase.

It is worth noting here that previous work found that the normal sequence of Pt silicidation is Pt → Pt_2Si → PtSi, and that 300 ℃ is insufficient to form PtSi [38, 39]. Pt_2Si forms first, and converts to PtSi upon sufficient supply of Pt atoms at high temperature (~ 700 ℃). It was also shown that a thin oxide layer at the Pt/Si interface can reduce Pt diffusivity, enabling Pt_2Si → PtSi conversion at temperature much lower than 700 ℃ [32]. Because both F and OI layers reduce Pt diffusion, there is insufficient supply of Pt atoms for PtSi formation in the F-implanted OI FGA sample as compared with the OI FGA sample.

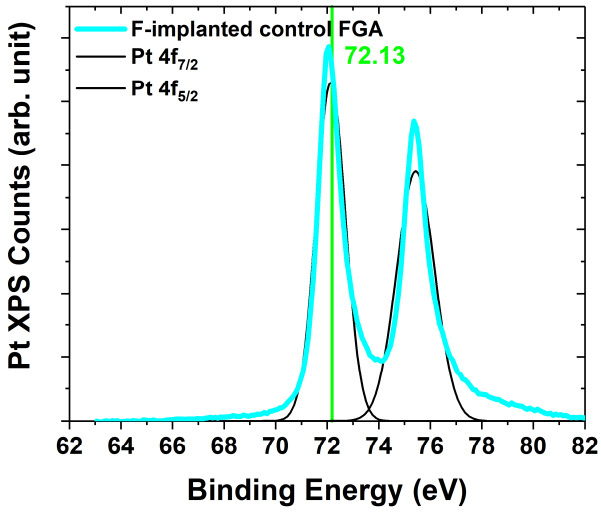

Fig. 5.18. XPS spectrum data of Pt 4f peaks for F-implanted control FGA sample.

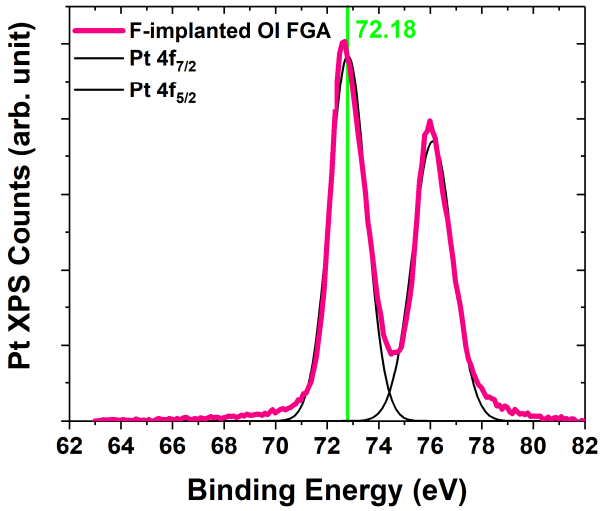

Fig. 5.19. XPS spectrum data of Pt 4f peaks for F-implanted OI FGA sample.

5.5 Summary

Both OI technology and F implantation are found to reduce Φ_{Bp} for a Pt/Ti/p-type Simetal-semiconductor contact because of enhanced Ti 2p core energy level shifts. OI technology is demonstrated to be more effective, reducing Φ_{Bp} by more than 50%, from 664 mV to 330 meV, after FGA. With OI technology, BF_2 implantation should be avoided for p-type S/D doping to achieve the lowest Φ_{Bp}.

5.6 References

[1] K. J. Kuhn, "Considerations for ultimate CMOS scaling," *IEEE transactions on Electron Devices,* vol. 59, no. 7, pp. 1813-1828, 2012.

[2] C. Lavoie *et al.*, "contacts in advanced CMOS: History and emerging challenges," *ECS Transactions,* vol. 77, no. 5, pp. 59-79, 2017.

[3] H. Niimi *et al.*, "Sub-$10^{-9}\,\Omega \cdot cm^{-2}$ n-Type Contact Resistivity for FinFET Technology," *IEEE Electron Device Letters,* vol. 37, no. 11, pp. 1371-1374, 2016.

[4] O. Gluschenkov *et al.*, "FinFET performance with Si: P and Ge: Group-III-Metal metastable contact trench alloys," in *2016 IEEE International Electron Devices Meeting (IEDM)*, 2016: IEEE, pp. 17.2. 1-17.2. 4.

[5] Y. Yang *et al.*, "Ultra low p-type SiGe contact resistance FinFETs with Ti silicide liner using cryogenic contact implantation amorphization and solid-phase epitaxial regrowth (SPER)," in *2016 IEEE Symposium on VLSI Technology*, 2016: IEEE, pp. 1-2.

[6] V. Kamineni *et al.*, "Tungsten and cobalt metallization: A material study for MOL local interconnects," in *2016 IEEE International Interconnect Technology Conference/Advanced Metallization Conference (IITC/AMC)*, 2016: IEEE, pp. 105-107.

[7] A. Agrawal *et al.*, "Fermi level depinning and contact resistivity reduction using a reduced titania interlayer in n-silicon metal-insulator-semiconductor ohmic contacts," *Applied Physics Letters,* vol. 104, no. 11, p. 112101, 2014.

[8] D. Connelly, C. Faulkner, P. Clifton, and D. Grupp, "Fermi-level depinning for low-barrier Schottky source/drain transistors," *Applied physics letters,* vol. 88, no. 1, p. 012105, 2006.

[9] J.-Y. J. Lin, A. M. Roy, A. Nainani, Y. Sun, and K. C. Saraswat, "Increase in current density for metal contacts to n-germanium by inserting TiO_2 interfacial layer to reduce Schottky barrier height," *Applied Physics Letters,* vol. 98, no. 9, p. 092113, 2011.

[10] R. Islam, G. Shine, and K. C. Saraswat, "Schottky barrier height reduction for holes by Fermi level depinning using metal/nickel oxide/silicon contacts," *Applied Physics Letters,* vol. 105, no. 18, p. 182103, 2014.

[11] J. Hu, K. C. Saraswat, and H.-S. P. Wong, "Metal/III-V Schottky barrier height tuning for the design of nonalloyed III-V field-effect transistor source/drain contacts," *Journal of Applied Physics,* vol. 107, no. 6, p. 063712, 2010.

[12] J. Padilha, A. Fazzio, and A. J. da Silva, "Van der Waals heterostructure of phosphorene and graphene: tuning the Schottky barrier and doping by electrostatic gating," *Physical review letters,* vol. 114, no. 6, p. 066803, 2015.

[13] J. D. Plummer, *Silicon VLSI technology: fundamentals, practice and modeling*. Pearson Education India, 2009.

[14] C. Ting and B. Crowder, "Electrical properties of Al/Ti contact metallurgy for VLSI application," *Journal of The Electrochemical Society,* vol. 129, no. 11, pp. 2590-2594, 1982.

[15] J. Kittl *et al.*, "Ni-and Co-based silicides for advanced CMOS applications," *Microelectronic Engineering,* vol. 70, no. 2-4, pp. 158-165, 2003.

[16] C.-N. Ni *et al.*, "Ultra-low contact resistivity with highly doped Si: P contact for nMOSFET," in *2015 Symposium on VLSI Technology (VLSI Technology)*, 2015: IEEE, pp. T118-T119.

[17] M.-A. Nicolet and S. Lau, "Formation and characterization of transition-metal silicides," in *VLSI Electronics Microstructure Science*, vol. 6: Elsevier, 1983, pp. 329-464.
[18] C. Lal, R. Dhunna, and I. Jain, "Phase formation, chemical composition and electrical studies of Ti/Si bilayer system," *Vacuum,* vol. 83, no. 6, pp. 931-935, 2009.
[19] F. Meli, Z. Sheng, I. Vedel, and L. Schlapbach, "XPS analysis of the getter mechanism and getter activation process," *vacuum,* vol. 41, no. 7-9, pp. 1938-1940, 1990.
[20] E. H. Rhoderick, "Metal-semiconductor contacts," *IEEE Proceedings I-Solid-State and Electron Devices,* vol. 129, no. 1, p. 1, 1982.
[21] E. Simoen, A. Rothschild, B. Vermang, J. Poortmans, and R. Mertens, "Impact of forming gas annealing and firing on the Al_2O_3/p-Si interface state spectrum," *Electrochemical and Solid-State Letters,* vol. 14, no. 9, pp. H362-H364, 2011.
[22] R. Carter *et al.*, "Passivation and interface state density of SiO_2/HfO_2-based/polycrystalline-Si gate stacks," *Applied physics letters,* vol. 83, no. 3, pp. 533-535, 2003.
[23] M. S. Arshad, Š. Trafela, K. Ž. Rožman, J. Kovač, P. Djinović, and A. Pintar, "Determination of Schottky barrier height and enhanced photoelectron generation in novel plasmonic immobilized multisegmented (Au/TiO_2) nanorod arrays (NRAs) suitable for solar energy conversion applications," *Journal of Materials Chemistry C,* vol. 5, no. 40, pp. 10509-10516, 2017.
[24] J.-S. Jang and T.-Y. Seong, "Mechanisms for the reduction of the Schottky barrier heights of high-quality nonalloyed Pt contacts on surface-treated p-GaN," *Journal of Applied Physics,* vol. 88, no. 5, pp. 3064-3066, 2000.
[25] P. Grunthaner, F. Grunthaner, and A. Madhukar, "An XPS study of silicon/noble metal interfaces: Bonding trends and correlations with the Schottky barrier heights," *Physica B+C,* vol. 117, pp. 831-833, 1983.
[26] J. Sun, K. Rickert, J. M. Redwing, A. Ellis, F. Himpsel, and T. Kuech, "p-GaN surface treatments for metal contacts," *Applied Physics Letters,* vol. 76, no. 4, pp. 415-417, 2000.
[27] M. Kuznetsov, J. F. Zhuravlev, and V. Gubanov, "XPS analysis of adsorption of oxygen molecules on the surface of Ti and TiN_x films in vacuum," *Journal of electron spectroscopy and related phenomena,* vol. 58, no. 3, pp. 169-176, 1992.
[28] Q. Wang and R. J. Madix, "Preparation and reactions of V_2O_5 supported on TiO_2 (1 1 0)," *Surface science,* vol. 474, no. 1-3, pp. L213-L216, 2001.
[29] Q. Sun, Y. Fu, J. Liu, A. Auroux, and J. Shen, "Structural, acidic and redox properties of V_2O_5-TiO_2-SO_{42}−catalysts," *Applied Catalysis A: General,* vol. 334, no. 1-2, pp. 26-34, 2008.
[30] L. Brillson, M. Slade, H. Richter, H. VanderPlas, and R. Fulks, "Titanium–silicon and silicon dioxide reactions controlled by low temperature rapid thermal annealing," *Journal of Vacuum Science & Technology A: Vacuum, Surfaces, and Films,* vol. 4, no. 3, pp. 993-997, 1986.
[31] C.-K. Lee, C.-D. Hsieh, and B.-H. Tseng, "Effects of a titanium interlayer on the formation of platinum silicides," *Thin Solid Films,* vol. 303, no. 1-2, pp. 232-237, 1997.
[32] C. Crider, J. Poate, J. Rowe, and T. Sheng, "Platinum silicide formation under ultrahigh vacuum and controlled impurity ambients," *Journal of Applied Physics,* vol. 52, no. 4, pp. 2860-2868, 1981.

[33] B. Y. Tsui and M. C. Chen, "Role of fluorine atoms on the thermal stability of the silicide/silicon structure," *Journal of applied physics,* vol. 76, no. 3, pp. 1995-1997, 1994.
[34] B.-Y. Tsui, J.-Y. Tsai, T.-S. Wu, and M.-C. Chen, "Effect of fluorine incorporation on the thermal stability of PtSi/Si structure," *IEEE Transactions on electron devices,* vol. 40, no. 1, pp. 54-63, 1993.
[35] A. Mokhberi, R. Kasnavi, P. B. Griffin, and J. D. Plummer, "Fluorine interaction with point defects, boron, and arsenic in ion-implanted Si," *Applied Physics Letters,* vol. 80, no. 19, pp. 3530-3532, 2002, doi: 1.1479458.
[36] X. Zhang *et al.*, "Effects of oxygen-inserted layers on diffusion of boron, phosphorus, and arsenic in silicon for ultra-shallow junction formation," *Journal of Applied Physics,* vol. 123, no. 12, 2018.
[37] X. Zhang *et al.*, "Effects of oxygen-inserted layers and oxide capping layer on dopant activation for the formation of ultrashallow pn junctions in silicon," *Journal of Vacuum Science & Technology B, Nanotechnology and Microelectronics: Materials, Processing, Measurement, and Phenomena,* vol. 36, no. 6, 2018.
[38] R. T. Fryer and R. J. Lad, "Synthesis and thermal stability of Pt_3Si, Pt_2Si, and PtSi films grown by e-beam co-evaporation," *Journal of Alloys and Compounds,* vol. 682, pp. 216-224, 2016.
[39] Y. Kumar, M. Yang, K. Ihm, K.-J. Lee, and C. C. Hwang, "Spatially-resolved spectroscopic investigations on the formation of PtSi by annealing on a silicon surface," *Chemical Physics Letters,* vol. 608, pp. 324-327, 2014.

Chapter 6

Conclusion

6.1 Summary

After five glorious decades, the exponential growth trend in IC complexity known as Moore's law is running out of steam, as shown in Figure 6.1. This is because the physical dimensions of transistors are approaching a fundamental limit, the size of an atom. In the "more than Moore" regime, the semiconductor industry needs to adopt novel technologies, employ new materials and physics to keep producing IC chips with improved transistor performance and more functionality. This work investigates the benefits of a CMOS performance booster "oxygen insertion" (OI) technology for improving advanced planar bulk-Si MOSFET and FinFET performances.

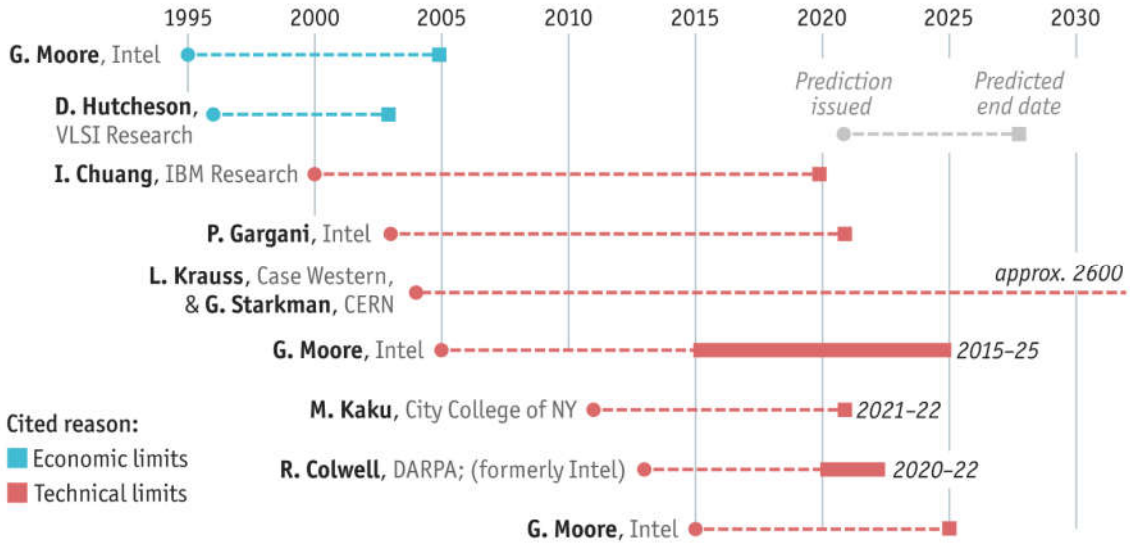

Fig. 6.1. Selected predictions for the end of Moore's law by semiconductor research experts (adapted from [1]).

The contributions of this dissertation can be summarized as follows:

Chapter 2 covers a comprehensive analysis of current research work on OI technology. Through experiments and TCAD simulations, OI technology is validated to provide simultaneous e- and h+ mobility enhancement while providing for reduced gate leakage currents. This is attributed to the local "quantum-confinement" effects produced by the OI layers. In addition, OI technology is experimentally verified to favorably retard boron and phosphorus transient-enhanced diffusion (TED) effects and produce a dopant pile-up effect. The physical mechanisms behind the improved doping profile is retarded interstitial-driven diffusion in the presence of the OI layers. These findings suggest that OI technology is a promising candidate as an advanced MOSFET performance booster.

Figure 6.2 shows that short-channel planar bulk MOSFET technology is predominantly used to produce semiconductor integrated circuits today. Leveraging the capability of OI technology to reduce TED effects, ultra-shallow doping junctions (USJ) suitable for the formation of p+/n and n+/p junctions are studied in **Chapter 3**. SIMS measurements reveal that OI layers are beneficial for reducing X_J by impeding the diffusion of Si interstitials, whereas neither a SiN_x capping layer nor a low-temperature oxide (LTO) capping layer helps to reduce X_J. Through the fabrication of R_{sh} test structures, electrical measurements, and TCAD simulations, OI technology is demonstrated to mitigate the increase in R_{sh} with X_J scaling thanks to its capability to enhance dose retention during thermal anneals. It is also demonstrated, for the first time, that an LTO capping layer causes an unfavorable increase in R_{sh} due to lower dopant activation levels. OI technology is shown to alleviate this detrimental effect. Overall, OI technology is verified to facilitate low-resistivity USJ formation for advanced planar MOSFETs for reduced X_J and lower R_{sh}.

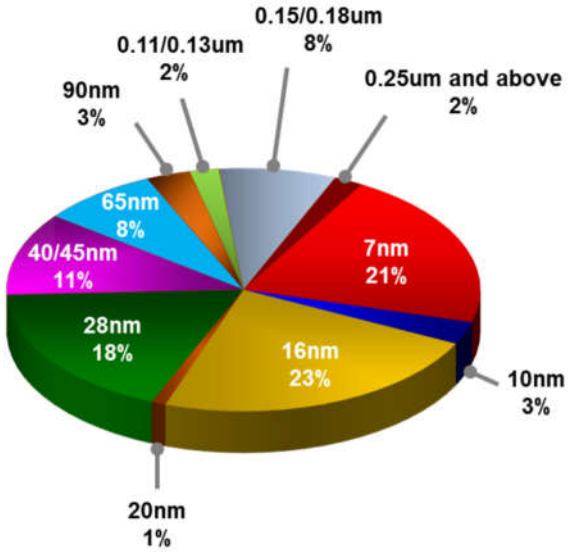

Fig. 6.2. 2019 second-quarter revenue by technology nodes (source: Taiwanese Semiconductor Manufacturing Company [2]).

Chapter 4 extends the evaluation of OI technology to advanced FinFET technology, targeting 7/8 nm low power technology node. Variability-induced transistor performance variations limit manufacturing yields and degrade IC performance due to mismatch. To choose the best candidate for 7/8 nm node, three FinFET technologies are considered: (1) the conventional bulk-Si (control) FinFET technology with a heavily-doped fin channel doping profile, (2) the bulk-Si (SSR) FinFET technology with a super-steep retrograde fin channel doping profile achievable with OI technology, (3) the SOI FinFET technology. TCAD simulation results demonstrate that SSR FinFETs can achieve higher I_{on}/I_{off} ratios and reduce the sensitivity of device performance to variations compared to control FinFETs. Then, 6T-SRAM bit cell performance and yield are estimated using a calibrated compact model. Leveraging the improved robustness against variability, both SSR FinFET and SOI FinFET technologies can reduce $V_{DD,min}$ by as much as 100 mV compared with the control FinFET technology. Due to the marginal improvement in mitigating transistor performance variations, SOI technology only provides for slightly smaller $V_{DD,min}$ for 1-2-2, and 1-3-3 bit cells, and same $V_{DD,min}$ for 1-1-1 bit cells. This study demonstrates that the SSR FinFET technology can be used as a cheaper alternative to the SOI FinFET technology for extending CMOS scaling beyond the 10-nm node.

With the continued miniaturization of transistors, contact resistance (R_{CON}) of metal/semiconductor (M/S) contacts has become a dominant parasitic resistance ($R_{parasitic}$) component [3, 4], which can degrade transistor performance by lowering g_m and $I_{D,sat}$ due to reduced V_{ov} ($= V_{GS} - V_{th}$). To overcome this challenge, it is desirable to fabricate M/S contacts that exhibit near-ideal ohmic contact behaviors [5-7].

Chapter 5 studies the effects of OI technology and fluorine (F) implantation to reduce the Schottky barrier height (Φ_{Bp}) of a Pt/Ti/p-type Si metal-semiconductor (M/S) contact. Through fabrication of Schottky diodes, and electrical measurements, OI technology is demonstrated to be more effective than F for Φ_{Bp} reduction, both before and after forming gas anneal (FGA). Secondary ion mass spectrometry (SIMS) and X-ray photoelectron spectroscopy (XPS) analyses were performed to gain insights into the mechanisms behind Φ_{Bp} reduction. For unannealed samples, OI layers facilitate Ti-O and Si-O bond formation, which can enhance Ti 2p and Si 2p core energy level shifts. FGA can reduce Φ_{Bp} by more than 50% (664 meV → 330 meV) with the presence of OI layers, whereas FGA only reduces Φ_{Bp} by 3.7% (563 meV → 542 meV) with the presence of F. XPS analyses of Pt $4f_{7/2}$ core level reveal that the Pt diffusivity during FGA is critical to Pt silicidation phase. Because OI layers can suppress Pt diffusion more effectively than F, OI layers promote low-Φ_{Bp} Pt monosilicide formation. It is found that the co-existence of OI layers and F almost block Pt diffusion completely and thus prevent Pt silicide formation. This indicates BF_2 implantation shall be avoided in the presence of OI technology. Overall, OI technology is demonstrated to be a more effective candidate for reducing p-type contact resistance.

6.2 Future Directions

Chapter 3 demonstrates that OI technology can suppress dopant diffusion and help retain dose during RTA treatment, facilitating low-resistivity ultra-shallow junction formations. It is worthwhile to experimentally verify the reduction in S/D parasitic resistances through the fabrication of short-channel planar bulk-Si MOSFETs. To augment the benefits of OI technology, it is desirable to optimize ion implantation parameters which have been fine-tuned for the baseline process.

With shrinking contact sizes, the MOSFET's external resistance (R_{EXT}) has become a significant component in the MOSFET's on-state resistance (R_{ON}). Much research is done to overcome this challenge in terms of (1) S/D epitaxial growth (2) contact/extension doping (3) middle of line (MOL) metallization [8]. Silicon germanium (SiGe) epitaxial S/D regions have been widely adopted to boost hole mobility for advanced pFETs [9, 10]. Incorporating partial oxygen monolayers into SiGe during epitaxy process may also help to achieve a higher level of dopant activation and improve dose retention during RTA treatments.

Chapter 4 demonstrates the benefits of SSR FinFETs achievable with OI technology for improved robustness against variations and higher I_{on}/I_{off} ratios as compared with control FinFETs. Recent research shows that the diffusion of phosphorus atoms from the S/D epi into the fin channel region during spike RTA can degrade SCE [11]. Because the OI layers used for SSR fin doping profile formation are incorporated around the fin bottom, it is worthwhile and natural to utilize them as a dopant diffusion barrier layer to reduce phosphorus diffusion.

In **Chapter 5**, OI layers and F are verified experimentally to reduce Φ_{Bp} of a Pt/Ti/p-type Si metal-semiconductor (M/S) contact. Recent research shows a trade-off exists between Φ_{Bp} and the active boron (B) concentration of SiGe S/D epi for advanced p-type FinFETs. A higher Ge concentration can help reduce Φ_{Bp} due to Fermi-level pinning near the valence band edge, but also reduces B solubility so the overall benefits diminish [8]. OI technology is promising to overcome this challenge by facilitating Ge-O bond formation, which may help achieve lower Φ_{Bp} at the same active B concentration as for the baseline process. However, to prevent channel mobility degradation, the impacts of OI layers on strain at various Ge concentrations should be investigated simultaneously.

6.3 References

[1] "After Moore's Law," in *The Economist*, ed. London, United Kingdom, 2016.

[2] E. Sun, "2019 Second Quarter Earnings Conference Presentation Materials," TSMC, 2019.
[3] S.-D. Kim, C.-M. Park, and J. C. Woo, "Advanced model and analysis of series resistance for CMOS scaling into nanometer regime. II. Quantitative analysis," *IEEE T Electron Dev,* vol. 49, no. 3, pp. 467-472, 2002.
[4] Z. Zhang *et al.*, "Sharp reduction of contact resistivities by effective Schottky barrier lowering with silicides as diffusion sources," *IEEE Electron Device Letters,* vol. 31, no. 7, pp. 731-733, 2010.
[5] K. Matsuzawa, K. Uchida, and A. Nishiyama, "A unified simulation of Schottky and ohmic contacts," *IEEE T Electron Dev,* vol. 47, no. 1, pp. 103-108, 2000.
[6] A. Agrawal *et al.*, "Fermi level depinning and contact resistivity reduction using a reduced titania interlayer in n-silicon metal-insulator-semiconductor ohmic contacts," *Applied Physics Letters,* vol. 104, no. 11, p. 112101, 2014.
[7] D. Connelly, C. Faulkner, P. Clifton, and D. Grupp, "Fermi-level depinning for low-barrier Schottky source/drain transistors," *Applied physics letters,* vol. 88, no. 1, p. 012105, 2006.
[8] H. Wu *et al.*, "Parasitic Resistance Reduction Strategies for Advanced CMOS FinFETs Beyond 7nm," in *2018 IEEE International Electron Devices Meeting (IEDM)*, 2018: IEEE, pp. 35.4. 1-35.4. 4.
[9] C. Auth *et al.*, "A 22nm high performance and low-power CMOS technology featuring fully-depleted tri-gate transistors, self-aligned contacts and high density MIM capacitors," in *2012 Symposium on VLSI Technology (VLSIT)*, 2012: IEEE, pp. 131-132.
[10] S. Natarajan *et al.*, "A 14nm logic technology featuring 2nd-generation FinFET, air-gapped interconnects, self-aligned double patterning and a 0.0588 μm^2 SRAM cell size," in *2014 IEEE International Electron Devices Meeting*, 2014: IEEE, pp. 3.7. 1-3.7. 3.
[11] S. Mochizuki *et al.*, "Advanced Arsenic Doped Epitaxial Growth for Source Drain Extension Formation in Scaled FinFET Devices," in *2018 IEEE International Electron Devices Meeting (IEDM)*, 2018: IEEE, pp. 35.2. 1-35.2. 4.