# FABRICATION, CHARACTERIZATION, AND MODELING OF SILICON MULTI-GATE DEVICES

ZHAO HUI

A THESIS SUBMITTED FOR THE DEGREE OF DOCTOR OF PHILOSOPHY
DEPARTMENT OF ELECTRICAL AND COMPUTER ENGINEERING
FACULTY OF ENGINEERING
NATIONAL UNIVERSITY OF SINGAPORE

*To Alick*

# EXECUTIVE SUMMARY

As multi-gate devices such as FinFET and nanowire FETs emerges as leading contenders of the future generation electron devices, detailed study of their electrical properties, characterization as well as effective modeling solution are much needed before they become truly viable for industrial application. This dissertation addresses the fabrication, characterization and modeling of silicon multi-gate transistors fabricated using the conventional CMOS platform. Its main purpose is to overcome some major challenges in both device fabrication and sub-femto farad capacitance measurement and modeling. A study of three dimensional electric field provided valuable insights to device operation and optimization for multi-gate devices. Charge Based Capacitance Measurement (CBCM) was simulated, analyzed, verified and applied for the first time to measurement of sub-femto farad voltage dependent capacitances. CBCM test keys were designed and fabricate for measurements of sub-femto scale nanowire capacitance. Also, measurement of charge and capacitance on single channel nanowire devices were used for self-consistent tight-binding computation of intrinsic and extrinsic capacitance calculation and extraction of series resistance and carrier mobility.

# ACKNOWLEDGEMENT

First and foremost I offer my most sincere gratitude to my supervisors, Prof. Ganesh S. Samudra and Dr. Subhash C. Rustagi, for their inspiration, encouragement, and guidance throughout this work. Without them this thesis would not have been completed or written. One simply could not wish for a better or friendlier supervisor and teacher.

My gratitude is also devoted to Dr. Patrick Lo, Dr. Navab Singh, and all other researchers, engineers and colleagues at Institute of Microelectronics, Singapore. They helped me in so many ways in completing my experiments. I feel extremely privileged to have been working with them.

I'm also grateful to Dr. Mark Lundstrom and his group at Purdue University for, collaboration and their insightful comments and discussions. Especially Mr. Raseong Kim and Mr. Abhijeet Paul, who has helped in carrying out the tight-binding calculation, in many ways I have learnt much from the both of them.

In my daily work I have been blessed with a friendly and cheerful group of fellow students. I thank Ms. Jiang Yu, Mr. Zang Hui, Mr. Wang Jian, Mr. Ma Fa-Jun, Ms. Fu Jia, Ms. Mao Shengchun for being such companiable classmates and faithful friends. I'm indebted to them for giving me so much help, support and encouragement. I'm also thankful to all students and engineers in SNDL who have worked together with me and brightened many a dull lunch time. I will cherish their friendship for a lifetime.

Finally, I thank my parents for their love and for supporting me through all my studies at Singapore.

I dedicate this work to my husband Alick whose love makes me go on.

# TABLE OF CONTENTS

# LIST OF FIGURES

# LIST OF TABLES

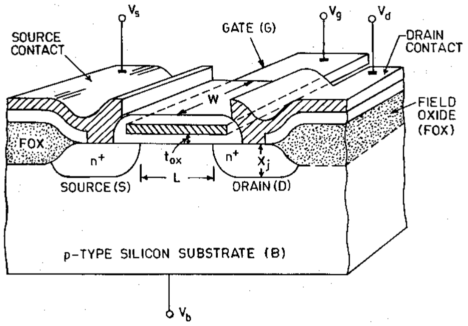
# Chapter 1

# Introduction

## 1.1 Planar MOSFET Scaling: History, Trends and Issues

The first commercially successful CPU product with gate length less than 100 nm were shipped by Intel© in year 2002, marking the end of Microelectronics Era and inaugurating the beginning of the Nanoelectronics Era [1]. In the course of less than 50 years, the semiconductor industry has grown from $1B of 1963 to more than $250B in 2007 [2] following essentially Moore's law [3] and guided by Dennard's scaling theory [4][5].

Dennard's scaling theory is based on device parameters and operating voltage scaling by the same factor, maintaining a constant electric field in the device. **Fig. 1-1** described the typical planar device structure and how the key technology parameters change according to Dennard's scaling law (κ is the scaling factor).



| Device Parameter | Scaling Factor |
|---|---|
| Device dimensions $t_{OX}$, L, W | 1/κ |
| Substrate doping $N_{sub}$ | κ |
| Voltage V | 1/κ |
| Current I | 1/κ |
| Capacitance $C_{ox}$ | 1/κ |
| Delay per gate VC/I | 1/κ |
| Power dissipation per circuit | $1/κ^2$ |

**Fig. 1-1**: Conventional planar MOSFET structure and constant field scaling theory

Moore and Dennard's theory set the technology on a course of regular developing pace, providing consistent improvements in transistor density, performance and power.

Each new generation of technology was expected to reduce minimum feature size by approximately 0.7x, which translates to a density improvement of 2x in every two years. To sustain the pace of scaling, numerous new technology and techniques has been developed and deployed overcoming the challenges of ever smaller device feature. Pocket and HALO implant, graded channel doping, and shallow junction, are but just a few of such examples. However, Moore himself has also noted, “no exponential is forever” [6]. Back in year 1998, the International Technology Roadmap for Semiconductor (ITRS) [7] has already identified that a new MOS transistor approach, departing from classical scaling and introducing new materials and properties, was necessary to continue to double the transistor density in every two years. For this purpose, silicide materials were used to reduce the series resistance in shallow junction devices. Strained silicon was adopted as a means of enhancing mobility without reducing doping. Subsequently, high-κ material based on hafnium and dual metal gate were also introduced, reducing gate leakage current and equivalent $SiO_2$ dielectric thickness at the same time without scaling down the physical thickness of the gate dielectric.

Looking forward, although the basic material for a MOSFET – silicon  and Si based compound material and silicides – are expected to lead commercialization of nanotechnology for the next two decades [2][8][9], there’s no doubt that scaling will be more challenging towards the end of the roadmap. The major ones of these challenges include controlling leakage currents and short-channel effects (SCEs), and maintaining control of device parameters such as threshold voltage despite the increasing variability in nanoscale.

### *1.1.1 Increasing dielectric field and short channel effects*

In order to boost the device performance, the gate dielectric thickness has been scaled more aggressively than the supply voltage, especially for the submicron technology nodes. For the past 20 years, the power supply only changes from 5 V for 1 μm technology to ~1.5V for the 0.1 nm technology generation. On the other hand, the gate oxide thickness has been scaled down from 200 Ǻ to less than 20 Ǻ (a few atomic layers thick) below which significant tunneling takes place. A direct result of this non-proportional scaling is that the oxide field has continually increased from 1 MV/cm to 5 MV/cm during the last decades, as shown in **Fig. 1-2**. There are two obvious advantage of increasing oxide filed:

(1) Neglecting the mobility degradation caused by higher field, the channel resistance is reduces by approximately the inverse of the field as in

$$R_{CHL} = \frac{L_{eff}}{W_{eff}\mu C_{ox}(V_{gs} - V_{TH})} \approx \frac{L_{eff}}{W_{eff}\mu\varepsilon_{si}E_{ox}} \qquad (1\text{-}1)$$

(2) Increasing oxide field improves the SCE by giving the gate better control of the channel potential, enabling more rapid scaling of the channel length.

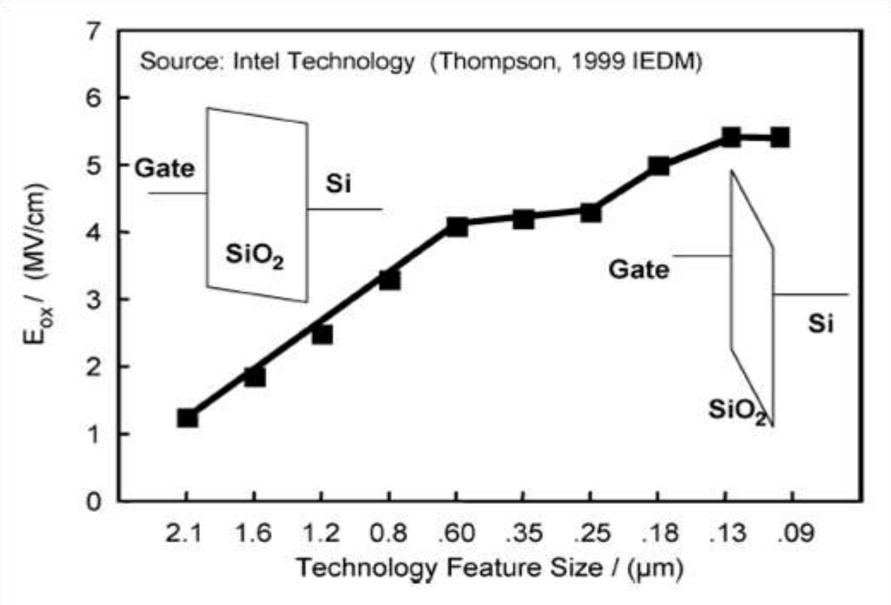


**Fig. 1-2**: Oxide field plotted against technology generations [10].

The minimum MOSFET channel length is ultimately set by the gate dielectric thickness relative the Si depletion width. As shown by Frank et al. [11], based on the 2D electrostatic and matching the boundary conditions of electric fields, a scale length λ can be derived as follows:

$$\varepsilon_{si} \tan(\pi t_i / \lambda) + \varepsilon_i \tan(\pi t_{si} / \lambda) = 0 \tag{1-2}$$

where $t_{si}$ is the depth of depletion region in silicon and $t_i$ is the thickness of the insulator. It was a shown that in the extreme limit, $\lambda \approx t_i + (\varepsilon_i / \varepsilon_{si}) / t_{si}$ , dominated by the physical thickness of the gate insulator. Therefore, there's a scaling limit set by the acceptable SCE for planar transistors even with high-κ gate insulators: the minimum useful channel length is about 1.5-2 times λ [12]. Empirically, it has been observed that for $V_{TH}$ and Drain-induced-barrier-lowering (DIBL) to be contained, the minimum lateral distance from source to drain is approximately 40 times the oxide thickness [13].

### *1.1.2 Leakage currents*

Aggressive gate dielectric thickness reduction causes higher gate leakage current. The minimum $SiO_2$ thickness is approximately 1.2 nm, or three to four atomic layers, resulting in a tunneling current of ~ 100 A/cm$^2$[14]. On the other hand, scaling of channel length and/or threshold voltage increases the subthreshold drain leakage due to DIBL. In some advanced technologies, the subthreshold leakage has become so high as to consume nearly quarter of the total power dissipated as shown in **Fig. 1-3**. Ultimately, the off-state leakage, due to the combination of gate and subthreshold leakage currents, will limit the smallest dielectric thickness and in turn the practical MOSFET channel length for planar MOSFETs.

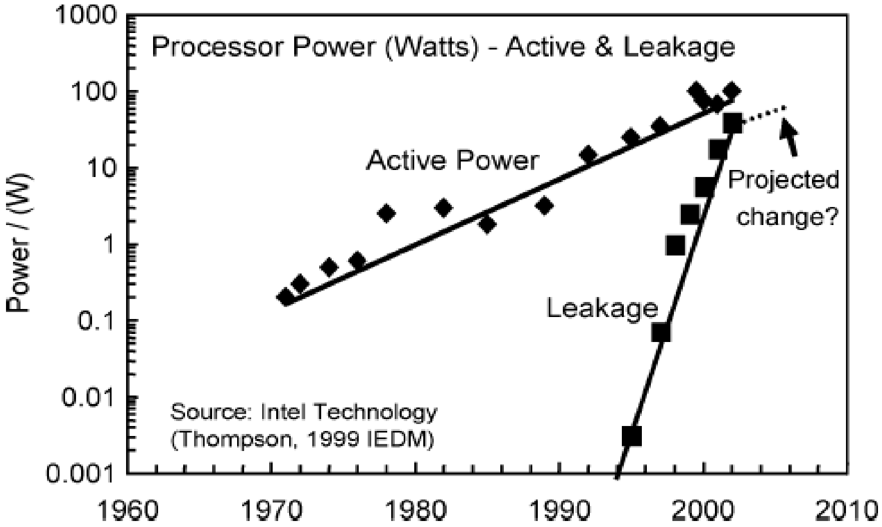


**Fig. 1-3**: Subthreshold leakage power plotted against year [13].

### *1.1.3 Variability*

As the device dimension scales into the submicron regime, the statistical variation of channel dopant number become more significant due to the smaller dopant population [15]. Increasing $V_{TH}$ shift was found to have detrimental effects in matching and device optimization [16]. On top of that, as the complexity of device increases and dimension decreases further, the variability in critical dimensions of the device will become increasingly significant [17][18]. Process induced variability due to uncertainties in lithography, etching and deposition will contribute to even greater dispersion of device parameters [18] as seen in some experimental nanowire devices [19].

## *1.2 Next Generation MOSFET Devices*

Addressing the three issues associated with scaling of the previous section without degradation of the device performance is not at all straightforward because of the trade-offs among three main indexes of MOSFWT performance: the on-current $I_{ON}$, the power consumption $P_{consum}$ and the SCEs. Schematic of such trade-off relationship is illustrated in **Fig. 1-4**. For low-power device application specifically, lower $V_{DD}$, higher threshold voltage $V_{th}$, higher substrate doping $N_{Sub}$ and higher electrical oxide thickness (EOT) is needed to reduce leakage current and stand-by power. However, all these are known to reduce $I_{ON}$ substantially and also hold inconsistency themselves.  Increasing oxide thickness $T_{OX}$, which is needed for reducing leakage current, increases sub-threshold swing (SS) for intrinsic gate capacitance $C_g$. And increasing $N_{Sub}$, which is necessary for suppressing SCE in bulk MOSFETs causes increase in leakage current on the account of junction tunneling current and GIDL. What's more, increasing $N_{Sub}$ also leads to degradation in mobility, reducing the $I_{ON}$ even further.

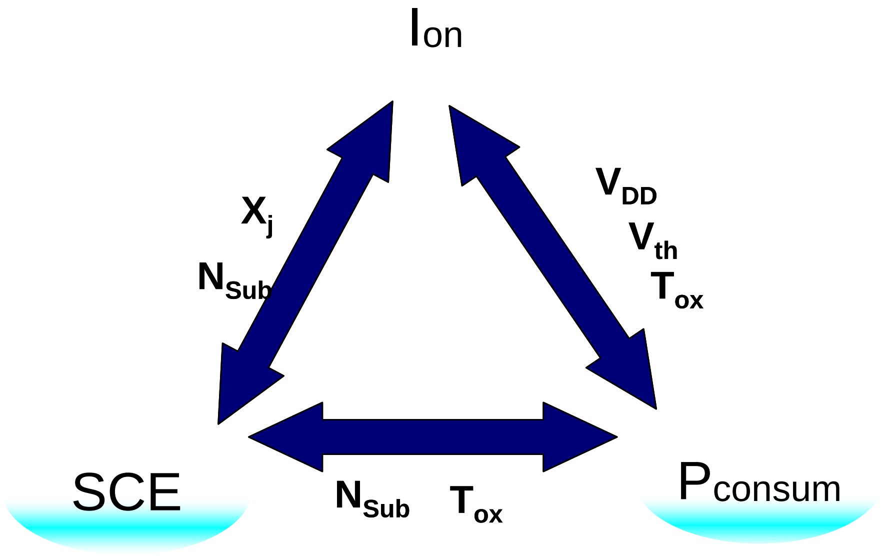


**Fig. 1-4**: The trade-off relationships among the three main indexes of performance: current drive $I_{ON}$, the SCE, and the power consumption $P_{consum}$. Listed along the arrows are the process/device parameters related to the three indexes.

The industrial and academic communities are pursuing two avenues to meet these challenges, namely, new materials and new transistor structures. New materials includes high-κ and metal gate for gate stack, channel material for enhanced mobility such as Ge and GaAs, as well as modified source/drain junction for improved resistance and carrier injection velocity. On the other hand, new Ultra-Thin-Body (UTB) transistor structures such as fully-depleted SOI FETs, nanowire (NW) FETs, carbon nanotube devices, and FinFETs, seek to improve the electrostatics of the MOSFETs, while providing a platform for integration of new materials for further performance enhancement at the same time. They are therefore widely regarded as the forerunners for scaling MOSFET devices towards the end of the roadmap.

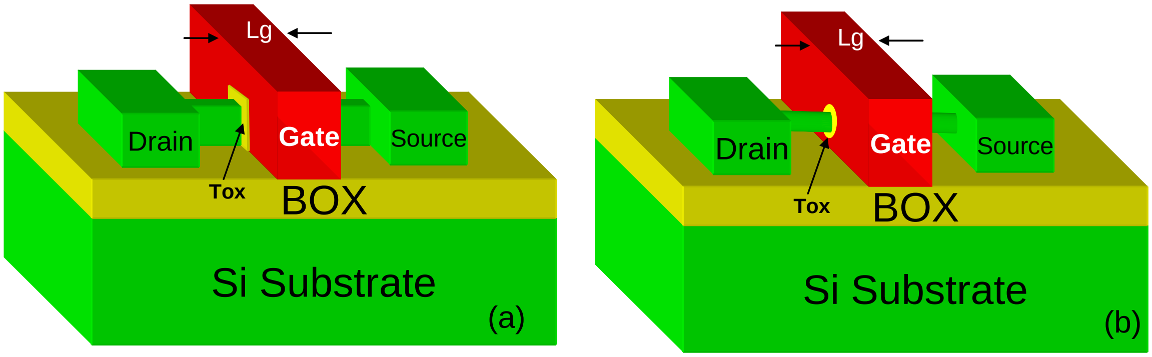


**Fig. 1-5**: Device structure illustrated for (a) – the tri-gate FinFET (b) – the nanowire FET

Two examples of the next generation device structures, the FinFET and the NW FET are shown in the schematics of **Fig. 1-5.** They both consists of a thin body in the shape of fin or wire with attached source and drain region for contact. The device was naturally isolated due to the SOI structure. The gate runs over the body forming a conducting channel. The gate width depends on the height and width of the fin in the case of FinFET; and on the circumference of the wire in the case of nanowire.

The key device concept behind most UTB device structures is improved SEC due to its fully-depleted nature and/or multiple gate structure. Better gate control results in nearly ideal subthreshold swing and reduced off-state current which in turn relaxed the requirement for $N_{sub}$ in scaling.  Power consumption can also be better optimized due to its much improved on and off current ratio. Therefore, technically the new UTB device structures, especially those with multiple and surround gate, have more potential in scaling than the conventional planar devices. However, in order to achieve the good SEC control, the body thickness in these devices needs to be considerably thinner than the gate length [20] which causes several issues:

(1) Higher extrinsic source/drain resistance: this is due to the thin body but also the difficulty in contacting them.

(2) Udoped body: ultra-thin body has to be undoped to maintain acceptable dopant fluctuation which may cause threshold voltage shifts. Complicated process such as dual metal process is needed to set matching threshold voltage for CMOS.

(3) Difficulty in characterization and modeling: the structures has been changed from 2D to 3D, therefore the old models for planar devices are no longer applicable. Adding to the complexity of 3D nanoscale geometry and electric field, is the possible change of band structure and other transport properties which only exist in nanoscale thin bodies [21].

## *1.3 Motivation of the Dissertation*

While the multi-gate UTB devices hold great promises in terms of device performance, breakthroughs in process technology, characterization technique and effective modeling solutions are needed before they can be viable.

The biggest challenge with process technology is the ability to produce thin Si channels with reasonable uniformity with the state-of-the-art CMOS technology platform [22]-[24]. Although the Extreme Ultraviolet (EUV) lithography system, due to be shipped commercially by 2010, will provide a better lithography solution to patterning finer lines, there is not yet an agreement on how the overall intra and inter-wafer uniformity for these devices can be best ensured [7]. Also, it is important to develop new process or material to reduce the source/drain junction resistance and contact resistance so that they do not negate the benefit of enhanced SCE control. Solutions such as raised S/D [25], fully-silicided (FUSI) S/D [30][31] has been proposed and implemented.

New characterization method needs to be developed to capture the ultra-low capacitance and charge associated with these multi-gate UTB devices [26][27]. The reality of variability in the nanoscale necessitate the measurements to be carried out on individual devices to investigate the detail of charge and transport. Here, overcoming the instrument resolution limitation and the low signal-to-noise ratio are the main challenges.

Finally, for evaluation of device and circuit performance, device models have to be redeveloped for the new 3D structures, taking into account the effects of 3D electric field distribution. Atomistic study of various bandstructure effects [21] in the electron transport of strongly quantized nanoscale channels [28][29] might be necessary and the results has to be validated before being incorporated into the new model.

This work seeks to study the nanoscale multi-channel device from fabrication, characterization and modeling perspective while focusing on some of the challenges described above. By extensive simulation, the scaling prospects of these devices were explored. Silicon based NW device were fabricated in conventional CMOS technology platform. Self-limiting oxidation [32] for producing nanometer scale channel with controllability were proposed and experimented. The devices were then characterized for I-V, C-V and mobility evaluation. Charge based capacitance measurement (CBCM) was studied in great detail with mix circuit and device mode simulations and applied for sub-femto-farad capacitance measurement in single channel NW devices. By the time of drafting this dissertation, this is perhaps the first successful measurement of single channel NW capacitance at room temperature. The C-V measurement results were also validated by three-dimensional electrostatic computations for parasitic estimation and two-dimensional self-consistent $sp^3s^*d^5$ tight-binding computations [33]-[35] for intrinsic gate capacitance calculations. The carefully designed CBCM technique thus emerges as a useful technique for measuring the capacitance and characterizing the transport in nano-scale devices. Although FinFETs and NW FETs are the subject of the study through most of the dissertation, the general technique used and conclusion derived is largely applicable to most of the devices in the “nanoscale multi-channel” family due to the similarities in their structures.

## 1.4 Outline of Dissertation

This dissertation is organized as follow:

The second chapter focuses on the scaling perspective of the nano-scale multi-gate transistor devices based on extensive process and device simulation study. Impact of fringing field, the dimension and material of spacer and the gate dielectric material are discussed in detail. Scaling is also discussed in a multi-channel context and the effect of channel pitch is delineated.

Chapter Three investigates the mechanism and accuracy of the technique specifically for sub-femto Farad nano-wire based device capacitance by carrying out extensive mixed device and circuit-mode simulations. The factors that influence the accuracy of the technique were identified.

Fabrication process for stringer-free NW device based on conventional CMOS technology platform will be discussed in detail in Chapter Four. Results of experimental and simulation study of self-limited oxidation process was also discussed.

In Chapter Five, the CBCM testkey design and fabrication were first introduced, followed by a demonstration of an innovative approach using CBCM to measure C-V and I-V characteristics from the same Silicon NW test structure. The measurement result of C-V measurement on single channel N and P-type NW FET will be presented and discussed.

Chapter Six reports on 2D and 3D analysis and modeling of single channel NW devices based on electrical characterization using CBCM technique. Self-consistent $sp^3s^*d^5$ tight-binding computations for intrinsic gate capacitance calculations and 3D simulation of extrinsic capacitance using COMSOL were shown to agree very well with experimental data. In addition, series resistance and mobility are extracted based on the experimental results.

An overall summary of this dissertation is presented in Chapter Seven. Key research contributions and suggestions for future research directions are highlighted.

## *Reference for Chapter 1*

[1] Intel© Website: http://www.intel.com/technology/

[2] P. A. Gargini, "The new scaling paradigm", *Proc. of VLSI-TSA 2008,* pp. 10-13

[3] G. E. Moore, "Progress in digital integrated electronics", *IEDM Tech. Dig.,* 1975, pp. 11-13

[4] R. Dennard, et al., "Design of ion-implanted MOSFETs with very small physical dimensions", *IEEE Journal of Solid State Circuits,* vol. SC-9, no. 5, pp. 256-268, 1974

[5] P. Chatterjee, "Device scaling: the treadmill that fueled three decades of semiconductor industry growth", *IEEE SSCS Newsletter,* vol. 12, no. 1, pp. 14-18, 2007

[6] G. E. Moore, "No exponential is forever", *ISSCC,* San Francisco, CA, 2003

[7] International Technology Roadmap for Semiconductors. http://www.itrs.net.

[8] R. Chau, S. Datta, M. Doczy, B. Doyle, B. Jin, J. Kavalieros, A. Majumder, M. Metz, and M. Radosavljevic, "Benchmarking nanotechnology for high performance and low power logic transistor applications", *IEEE Trans. On Nanotechnology,* vol. 4, no, 2, pp.153-158, 2005.

[9] T. Skotnicki, J. Hutchby, T.-J. King, H.-S. P. Wong, and F. Boeuf, "The end of CMOS scaling: towards the introduction of new materials and structural changes to improve MOSFET performance", *IEEE Circuits and Devices Magazine,* vol. 21, no. 1, pp. 16-26, 2005

[10] S. E. Thompson, "IEDM Short course on sub-100 nm CMOS", *IEDM 1999.*

[11] D. J. Frank, Y. Taur, and H.-S. Wong, “Generalized scale length for two-dimensional effects in MOSFETs” *IEEE Electron Device Lett.* vol. 19, no. 10, pp. 385-387, 1998.

[12] Y. Taur, “CMOS scaling beyond 0.1 μm: How far can it go?” *Proc. Of VLSI-TSA 1999,* pp. 6-9.

[13] S. E. Thompson, R. S. Chau, T. Ghani, K. Mistry, S. Tyagi, and M. Bohr, “In search of ‘forever,’ continued transistor scaling one new material at a time”, *IEEE Trans. on Semiconductor Manufacturing,* vol, 18, no.1, pp.26-36, 2005.

[14] S. Thompson *et al.,* “A 90 nm logic technology futuring 50 nm strained silicon channel transistors, 7 layers of Cu interconnects, low k ILD and 1 $\mu m^2$ SRAM cell ”, *Tech. Dig. IEDM 2002,* pp. 61-64, 2002.

[15] T. Mizuno, J. Okamura, and A. Toriumi, “Experimental study of threshold voltage fluctuation due to statistical variation of channel dopant number in MOSFET’s”, *IEEE Trans. Electron Devices,* vol. 41, no. 11, pp. 2216-2221 , 1994.

[16] P. A. Stolk, F. P. Widdershoven, and D. B. M. Klaassen, “Modeling statistical dopant fluctuation in MOS transistors”, *IEEE Trans. Electron Devices,* vol. 45, no. 9, pp. 1960-197, 1998.

[17] R. Wang, J. Zhuge, R, Huang, Y. Tian, H. Xiao, L. Zhang, C. Li, X. Zhang, and Y. Wang, “Analog/RF Performance of Si Nanowire MOSFETs and the Impact of Process Variation”, *IEEE. Trans. Electron Devices,* vol. 46, no. 6, pp. 1288-1294, 2007.

[18] Y.-S. Wu, and P. Su, "Sensitivity of gate-all-around nanowire MOSFETs to process variations – a comparison with multigate MOSFETs", *IEEE Trans. Electron Devices,* vol. 55, no. 11, 2008.

[19] S. D. Suk, *et al.,* "Investigation of nanowire size dependency on TSNWFET", *Tech. Dig. IEDM 2007,* pp. 891-894, 2007.

[20] J. G. Fossum, M. M. Chowdhury, V. P. Trivedi, T. J. King, Y. K. Choi, J. An and B. Yu, "Physical insights on design and modeling of nanoscale FinFETs", *IEDM Tech. Dig. 2003,* pp. 679-682, 2003.

[21] N. Neophytou, A. Paul, M. S. Lundstrom, G. Klimeck, "Bandstructure Effects in Silicon Nanowire Electron Transport" *IEEE Transactions on Electron Devices,* vol.55, no.6, pp.1286-1297, June 2008

[22] N. Singh, K. D. Buddharaju, S. K. Manhas, A. Agarwal, S. C. Rustagi, G. Q. Lo, N. Balasubramanian and D.-L. Kwong, "Si, SiGe nanowire devices by top-down technology and their applications", *IEEE Trans. on Electron Devices,* vol. 55, no. 11, pp. 3107-3118, 2008.

[23] V. Pott, K. E. Moselund, D. Bouvet, L. De Michielis, and A. M. Ionescu, "Fabrication and characterization of gate-all-around silicon nanowires on bulk silicon", *IEEE Trans. on Nanotech.,* vol. 7, no. 6, pp. 733-744, 2008.

[24] V. Subramanian, J. Kedzierski, N, Lindert, H. Tam, Y. Su, J. McHale, K. Cao, T. J. King, J. Bokor, and C. Hu, "A bulk-Si-compatible ultrathin-body SOI technology for sub-100 nm MOSFETs", *Proc. 1999 IEEE 57th Annual Device Research Conference Dig.,* pp. 28-29, 1999.

[25] Y. K. Choi, D. W. Ha, T. J. King, and C. M. Hu, “Nanoscale ultrathin body PMOSFETs with raised selective germanium source/drain”, *IEEE Electron Device Lett.,*vol. 22, no. 9, pp. 447-448, 2001.

[26] H. Zhao, S. C. Rustagi, N. Singh, F.-J. Ma, G. S. Samudra, K. D. Budhaaraju, S. K. Manhas,C.H. Tung, G. Q. Lo, G. Baccarani, and D. L. Kwong, “Sub-femto-farad capacitance-voltage characteristics of single channel gate-all-around nano wire transistors for electrical characterization of carrier transport”, *IEDM 2008 Tech. Dig.* pp.769-772, 2008

[27] R. Tu, Li Zhang, Y. Nishi, and H. Dai, “Measuring the Capacitance of Individual Semiconductor Nanowires for Carrier Mobility Assessment”, *Nano Lett.,* 7-6, pp1561-1565, 2007.

[28] T. Hiramoto, K. Miyaji and M. Kobayashi, “Transport in silicon nanowire and single electron transistors”, *IEEE Tech. Dig. Of Simulation of Semiconductor Process and Devices (SISPAD) 2007,* pp. 209-215, 2007.

[29] H. Minari and N. Mori, “Crystal orientation effects on ballistic hole current in ultrathin DG SOI MOSFETs”, *IEEE Tech. Dig. Of Simulation of Semiconductor Process and Devices (SISPAD) 2007,* pp. 229-232, 2007.

[30] E. T. Tan, K. L. Pey, N. Singh, G. Q. Lo, D. Z. Chi, Y. K. Chin, L. J. Tang, P. S. Lee and C. F. K. Ho, “Nickle-silicided schottky junction CMOS transistors with gate-all-around nanowire channels”, *IEEE Electron Device Lett.* Vol. 29, no. 8, pp. 902-905, 2008.

[31] E. T. Tan, K. L. Pey, N. Singh, G. Q. Lo, D. Z. Chi, Y. K. Chin, K. M. Hoe, G. Cui, and P. S. Lee, “Demonstration of Schottky barrier NMOS transistor with

Erbium silicided source/drain and silicon nanowire channel”, *IEEE Electron Device Lett.* Vol. 29, no. 10, pp. 1167-1170, 2008.

[32] H. I. Liu, D. K. Biegelsen, R. F. W. Pease, N. M. Johnson and F. A. Ponce, "Self-limiting oxidation of Si Nanowires", *J. Vac. Sci. Technol. B*, vol.11, no.6, pp. 2532-2537, 1993.

[33] G. Klimeck, F. Oyafuso, T. B. Boykin, R. C. Bowen, and P. von Allmen, "Development of a Nanoelectronic 3-D (NEMO 3-D) Simulator for Multimillion Atom Simulations and Its Application to Alloyed Quantum Dots" (invited), Computer Modeling in Engineering and Science (CMES), vol 3, no. 5, pp 601-642, 2002.

[34] Paul, Abhijeet; Luisier, Mathieu; Neophytou, Neophytos; Kim, Raseong; McLennan, Michael; Lundstrom, Mark; Klimeck, Gerhard (2006), "Band Structure Lab," doi: 10254/nanohub-r1308.6Y.

[35] N. Neophytou, A. Paul, M. S. Lundstrom, G. Klimeck, "Bandstructure Effects in Silicon Nanowire Electron Transport," IEEE Transactions on Electron Devices, vol.55, no.6, pp.1286-1297, June 2008.

# Chapter 2

# Simulation of FinFET and Their Scaling Properties

## *2.1 Introduction*

While standard planar device structure apparently faces insurmountable challenges in the sub-32nm low-power regime in terms of SCE and off-state leakage currents, ultra-thin body silicon-on-insulator (SOI) and FinFET devices are considered strong contenders for continuing scaling beyond 32 nm node. Three-dimensional (3D) nanoscale device structures, such as FinFET and Nano-wire (NW) devices in particular, offer not only excellent gate control, immunity to SCE but also much smaller junction capacitance. Compatibility with conventional CMOS fabrication process is also one of their major advantages. Therefore much research attention has been devoted to developing FinFET and NW technologies for next generation of low-power devices. On the other hand, for the sub-32nm FinFET devices reported thus far, large drive current and high switching speed, as projected by the International Technology Roadmap for Semiconductors (ITRS) [1] has not yet been observed [2]-[6]. Significant scattering of carriers [7]-[9], increased parasitic resistance due to thin and narrow Source/Drain (S/D) regions and increase in capacitance due to fringing fields in very small 3D geometry [10] are speculated to have an adverse impact on the performance of these devices.

Ever since the device gate dimension has been scaled into the sub-micrometer regime, parasitic capacitances have become a significant part of the gate capacitance and must be accounted for in evaluating device. They are also found to increase much faster as the scaling continues [10]-[12]. Trivedi et al. [11] first reported the effect of gate fringing field in Double-gate MOSFET on $C_{gg}$ using numerical simulation while discussing the effect of abrupt and under lapped gate profile. Bansal et al. [12] investigated the effect of fringing field component from gate sidewall to source through spacer in Double-gate MOSFET using conformal mapping. Both efforts point out the significant role parasitic play in sub-50nm transistor. However, both are based on a two-dimensional (2D) model the accuracy of which remain susceptible to error for most Fin-type transistors including structural variations such as Pi-gate MOSFET [13], Nanowire MOSFET [14] and gate-all-around MOSFET [15]. To fully capture the effect of fringing field on FinFET device performance, the non-planar nature of the device geometry and 3D fringing fields have to be taken into account. It is clear that the optimization of FinFET devices for adaptation in manufacture-worthy integrated circuits will require a thorough understanding, evaluation and optimization of these parasitic effects in 3D. In this chapter, we take FinFET as the subject of study. However, the methodology and general conclusion can be implied to a wider range of nanoscale devices such as OmegaFET and nanowire device which is typically a FinFET of much reduced in dimension with a gate-all-around structure.

The chapter is organized as follows. The 3D process and device simulation and the simulated device structure are described in section 2.2. The impact of fringing field is discussed in section 2.3 including the effects of these fields on the drive-current and off-state leakage for different spacer/gate electrode geometry. High-κ gate dielectric and merit of its physical thickness scaling are investigated in section 2.4 while section 2.5 describes the fringing fields in case of multiple fin structures with a view to investigate the role of fin-pitch. Section 2.6 discusses the implication of this work to other nanoscale transistor devices such as nanowire device. Section 2.7 summarizes and concludes the chapter.

## 2.2 The Device Structure, and, 3D Process and Device Simulation

### 2.2.1 Necessity of 3D device structure

Building a 3D process and device model is essential and necessary for sensible simulation study of FinFETs. As shown in **Fig. 2-1**, FinFETs are typically thin-body SOI devices having two vertical gates on two sides of the fin and a horizontal one on top of the fin. Therefore the three dimensional electric field must be considered for evaluation of capacitances in short channel FinFET devices. In most of 2D simulations, it was assumed that the structure extends in the vertical direction substantially more than the gate length which is in general not true for FinFETs. Simplifying the 3D field by combining 2D simulations of cross sections would inevitably lead to inaccuracy as a 2D model would fail to account for interactions with parts of the structure outside of the modeled plane. As shown in the $C_{gs}$ plot in **Fig. 2-2**, 2D simulation clearly over predicts the gate to source capacitance in FinFET by ignoring the underlying bottom oxide layer and assuming the device extends infinitely in the z-direction in solutions.

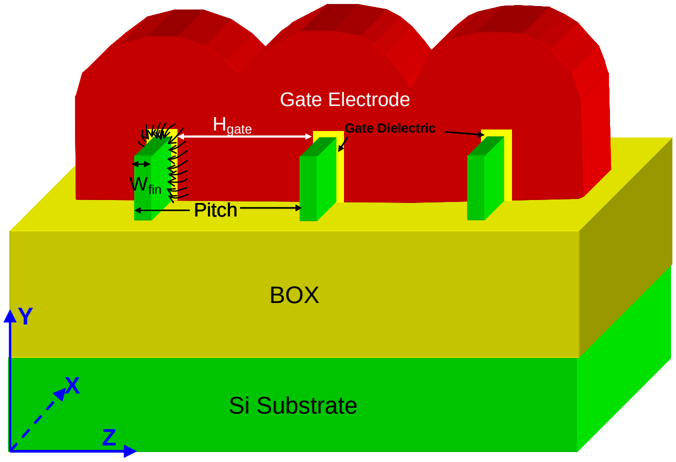


**Fig. 2-1**: (a) Three-dimensional (3D) schematic of a multiple-fin FinFET device showing the electric field liens from the gate to the source on the first fin. The spacer region and the raised S/D regions are not shown.

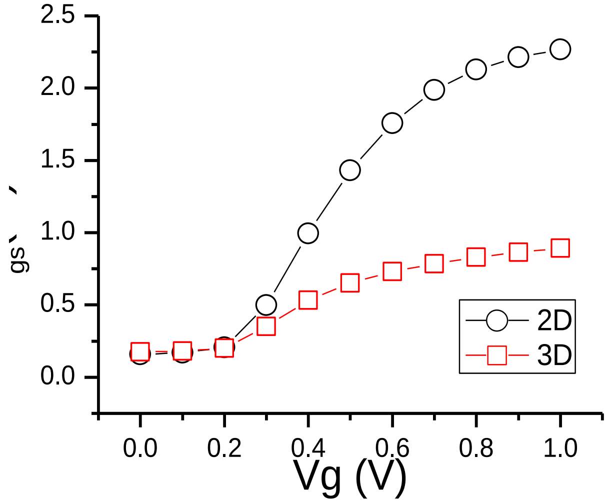


**Fig. 2-2**: Comparing the 3D simulation result of $C_{gs}$ of a FinFET with the 2D estimation (calculated based on average capacitance per unit gate length)

### 2.2.2 FinFET device with under-lapped gate structure

The FinFET devices with ultra narrow channels have been previously reported to relax the requirement of heavy channel doping for control of SCE, thus reducing the effect of impurity scattering on the effective mobility. While the body of FinFET is largely undoped to avoid threshold voltage shifting due to dopant fluctuation, part of the narrow S/D extension region connecting channel to the S/D heavily doped region are often designed to be undoped as well, forming an *under-lapped gate* [11] structure. Gate under lapping results in a bias-dependent effective channel length: The channel is longer in weak inversion and approaches the gate length in strong inversion.  Consequently, the SCE can be better suppressed while current drivability is maintained. It has been shown that by careful source/drain doping profile engineering, gate underlap structure in FinFET could reduce the subthreshold leakage by a few orders without sacrificing much on-state current [16]-[18]. However, this design leaves the portion of lightly doped channel under the under-lapping spacer weakly controlled by the gate and makes it sensitive to the fringing field from gate through the spacer due to its proximity. As a result, the thickness of the gate dielectric, geometry of the spacer and gate thickness show significant effects on channel electrical potential distribution through fringing field as shown in **Fig. 2-3**.

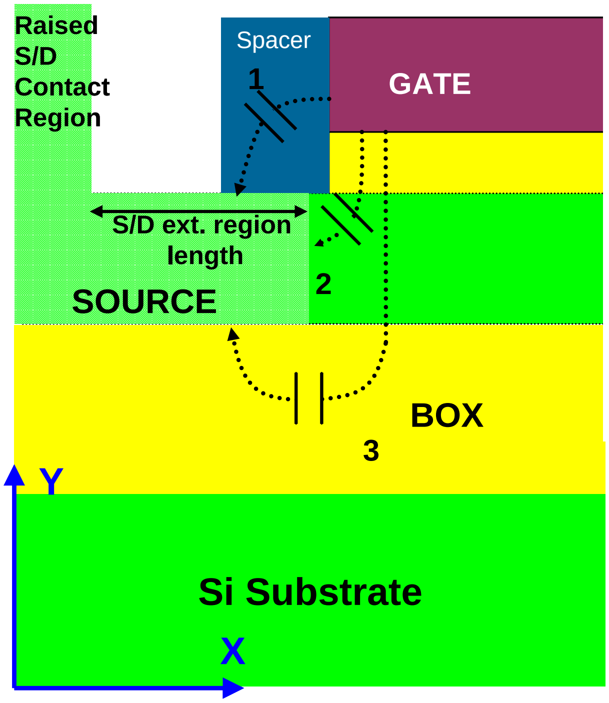


**Fig. 2-3**: A cross-section of the FinFET device along the cut line in Fig. 2-1. Three major components of the gate-to-source capacitance $C_{gs}$ are illustrated

### 2.2.3 3D process and device simulation and calibration

N-channel FinFET were fabricated on SOI substrate using 193 nm lithography systems. Fins were dry etched and annealed in $H_2$ ambient. $HfO_2$ gate dielectric was used with a $SiO_2$ liner layer. After a novel age stack [19][20]consisting of a combination of TiN and TaN and $SiO_2$ capping layer was deposited, two Arsenic implantations and one Phosphorous implantation were used for the source/drain doping and LDD formation [21][22]. For some devices, a selective epitaxial growth of Si was carried out on the HDD region for reduced series resistance. This process was simulated in Taurus Process [23] 3D process simulator and the simulated device structure (without raised HDD) as shown in **Fig. 2-4 (a).**

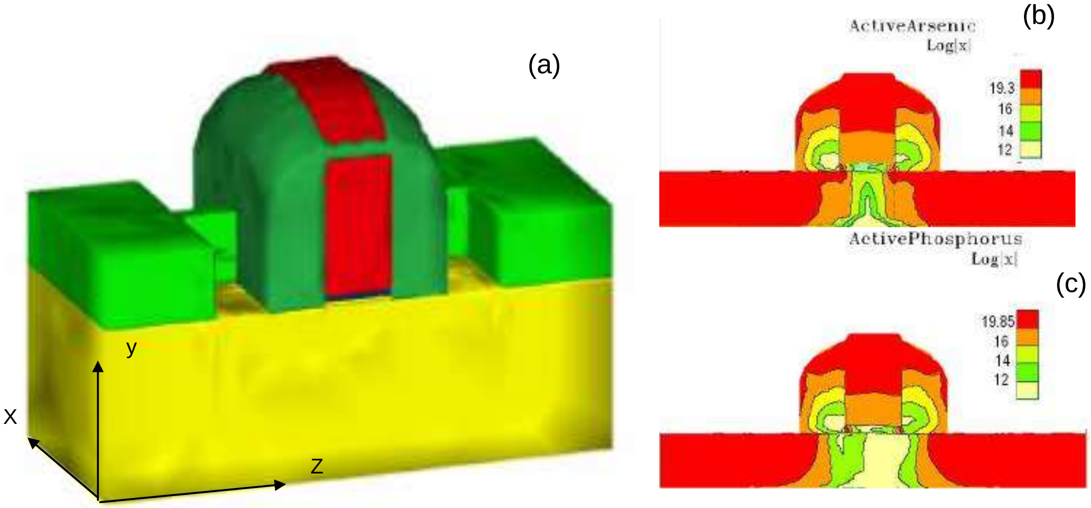


**Fig. 2-4**: (a) – FinFET structure simulated by Taurus Process™ 3D. (b)&(c) – the implant profile simulated shown in the device cross section for Arsenic and Phosphorus, respectively

For carrying out the 3D device simulations with computation efficiency, the source/drain region impurity profiles simulated by 3D process simulation as shown in **Fig. 2-4** (b) are represented by an analytical Gaussian function fitted to the process simulation data obtained from Taurus Process [24]. The work function and dielectric thickness was extracted by matching the gate capacitance of a sample fabricated device. An excellent fit was achieved as shown in **Fig. 2-5**. The doping profile was then fine tuned iteratively by fitting the simulated drain and gate electrical characteristics with experiment results. The device simulation included quantum effect, advanced mobility and recombination models [23]. $I_d$-$V_g$ curves both simulated and measured are shown in **Fig. 2-6**. The device properties such as gate length and dielectric thickness were then scaled according to ITRS guidelines for further study of FinFET devices.

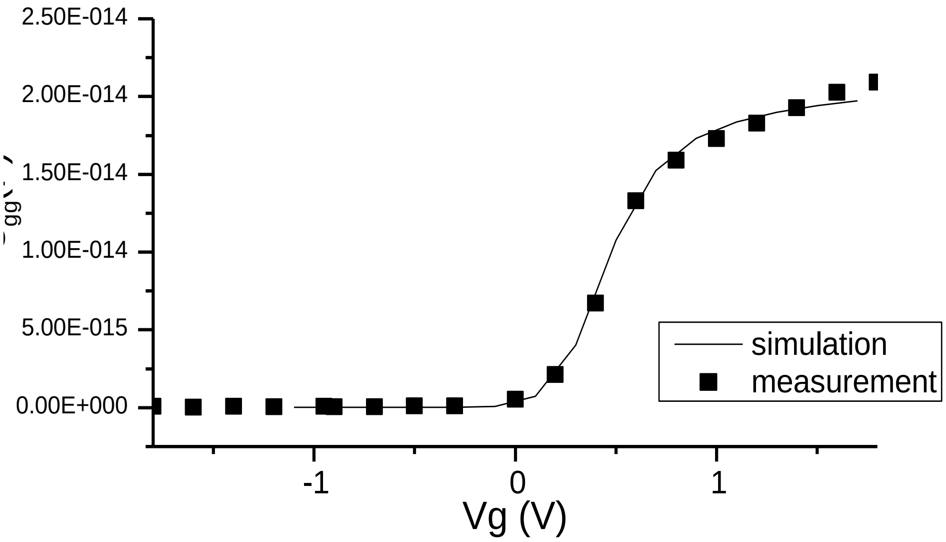


**Fig.2-5**: Gate work function and dielectric thickness fitting using C-V measurement data

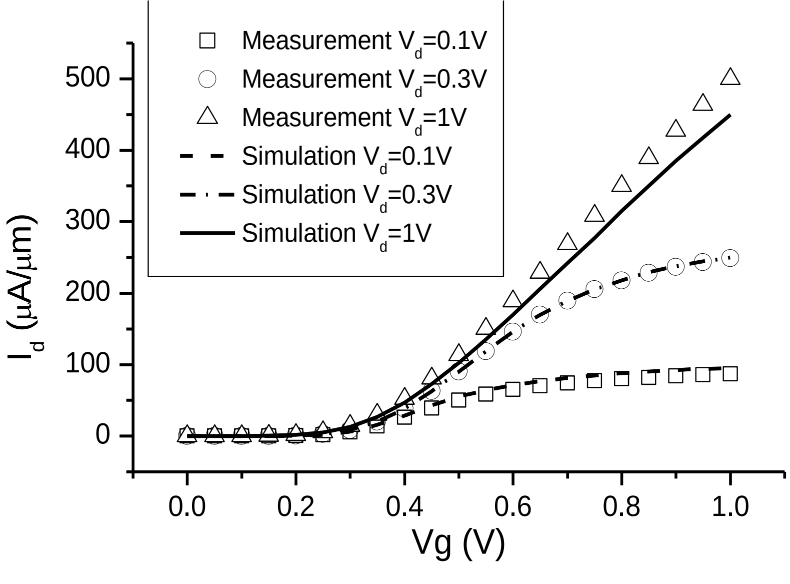


**Fig. 2-6**: Calibrated device gate characteristic

## 2.3 Effect of Fringing Field in Multiple Gate FinFET

Using the calibrated device as a reference, we constructed Low Stand-by Power Devices at 32 nm node in Taurus Device [23] for the rest of the study. The device has a physical gate length of 13 nm (with under lapping gate, the $L_{eff}$ is 18 nm), $V_{dd}$ of 0.9 V and 13 nm fin. A 100 nm thick bottom oxide (BOX) and 100 nm Si back substrate are also included to take into account the effect of the electric field terminating into the bulk on various capacitances [24]. The spacer width is taken as a variable which can be achieved in practice using post-implant trimming and other techniques [25][26], and, thus, the doping profile and the effective gate length in these simulations are assumed to remain unaffected by the change of the spacer width. All dielectric layers are assumed to be perfectly insulating so that leakage/tunneling current through dielectric materials is decoupled from S/D leakage.

Since the most basic and important device performance aspects are the current drive capability and switching speed, the major figures of merit considered in this work are $I_{ON}$ and $I_{OFF}$ and the intrinsic gate delay ($C_{gg}$•$V_{gs}$/$I_{ON}$). Here we define $C_{gg}$, “total effective gate capacitance” as the capacitive load represented by the FET to a driver device such as an inverter. In this case we have taken the load capacitance as the gate capacitance when both gate and drain voltages are at $V_{dd}$ relative to the source. $C_{gg}$ includes gate-to-source capacitance $C_{gs}$, gate-to-drain capacitance $C_{gd}$, and gate-to-substrate capacitance $C_{gb}$, with $C_{gs}$ being the dominating component.

Major components of the gate-to-source capacitances in a thin body FinFET device shown in **Fig. 2-3** are:

(1) Sidewall fringing capacitance from the sides of the gate to the source (also shown in 3D in **Fig. 2-1**);

(2) Fringing capacitance from the bottom of the gate to the source through the Si body; and

(3) Bottom-fringing capacitance from the gate to the source through the BOX layer.

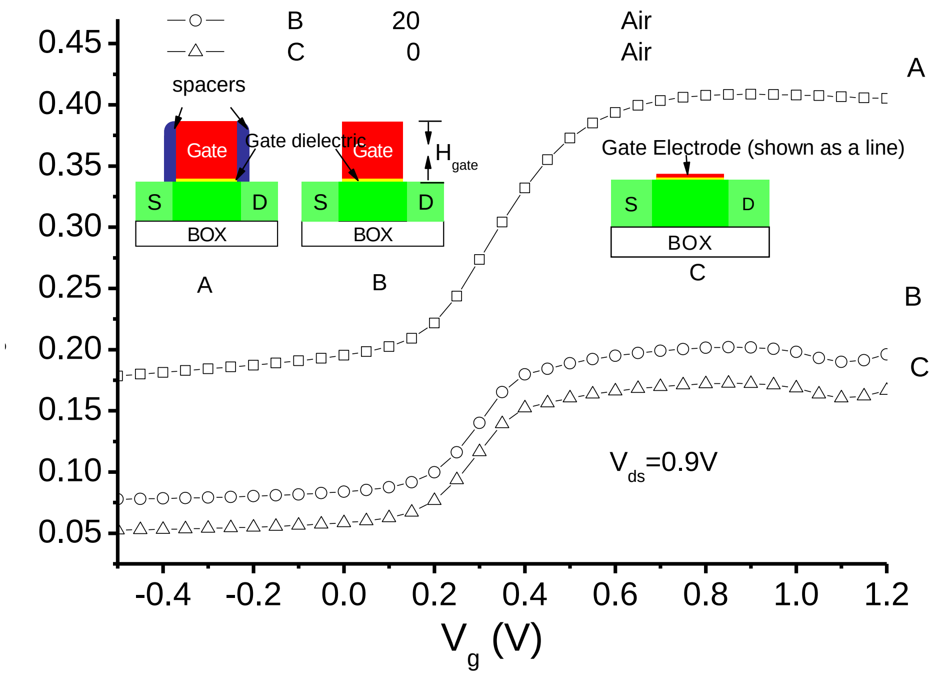


**Fig. 2-7**: Gate-to-source capacitances of three identical devices with different gate and spacer geometries. Device A- with volume gate electrode and nitride spacer. Device B- with volume gate electrode and no spacer. Device C- with surface gate electrode and no spacer. Inset: schematic cross-sections of the three FinFET devices.

We first consider three FinFET structures, named as A, B, and C, which are identical except for the gate and/or the spacer geometry with their schematics shown in inset of **Fig. 2-7**. Device A has a gate electrode thickness of 20 nm with a flanking nitride spacer of 20 nm thickness. Device B has the same gate electrode but without a spacer whereas device C has only a surface gate electrode (i.e. the thickness is negligible) of the same workfunction and no spacer. The significance of device structure to parasitic capacitance is reflected in **Fig. 2-8**, in which the gate-to-source capacitance $C_{gs}$ for three FinFET devices at a drain bias $V_{ds}$=0.9 V is plotted against $V_{gs}$. We are considering the device designed with an under lapped gate and thus the overlap capacitance is negligible here. The $C_{gs}$ of device A is found to be three times higher than that of device B and 3.5x higher than that of the device C in strong inversion. This difference among the devices is mainly due to capacitive coupling between the gate sidewall and the source [as represented by fringing capacitance component (1) in **Fig. 2-3**], which is significantly reduced when spacer material has lower permittivity (air in case of device B). Fringing capacitance is absent in case of a surface gate electrode (structure C). **Fig. 2-8** clearly quantifies that a significantly large fringing field that extends from the sides of the gate to the source through the spacer dominates the total gate-to-source capacitance.

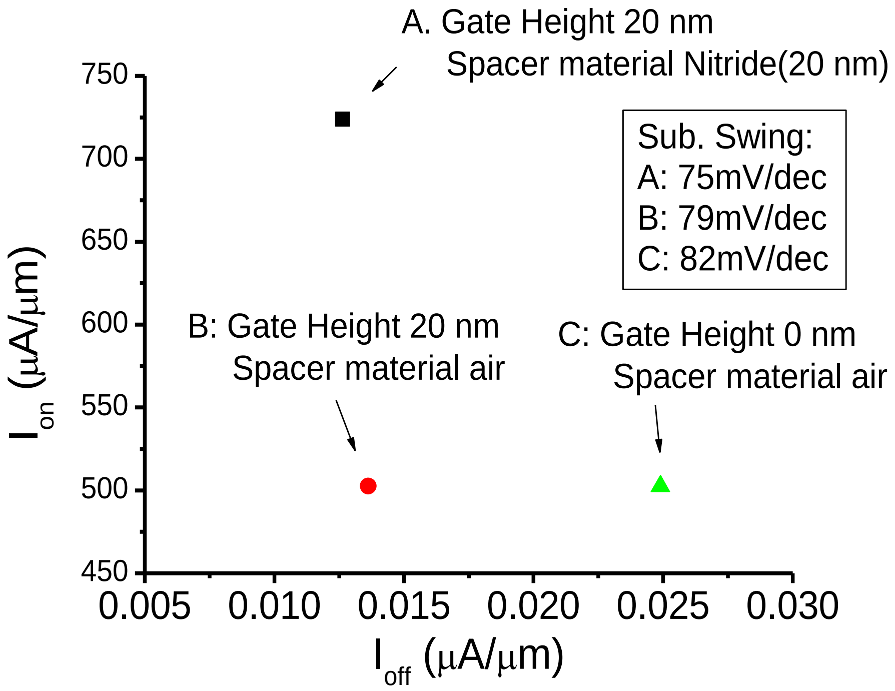


**Fig. 2-8**: Comparison of On- and Off-current of the three devices. A- with gate electrode height 20 nm and nitride spacer. B- with gate electrode height 20 nm and no spacer. C- with surface gate electrode height 0 nm and no spacer.

For short gate length device, the fringing field affects not only the capacitances; it also has modulating effect in the channel electrostatics because the two gate edges are in close proximity. As shown in **Fig. 2-8,** the structure of the spacer has a significant impact on the ON and OFF-state current. Device A, with a similar $I_{OFF}$ as that of device B, has $I_{ON}$ enhanced by 44%. While device C, with similar low $I_{ON}$ as device B, has 80% larger $I_{OFF}$ compared with that of B as predicted by 3D simulation. The effect of spacer fringing field will become clearer with further discussion below.

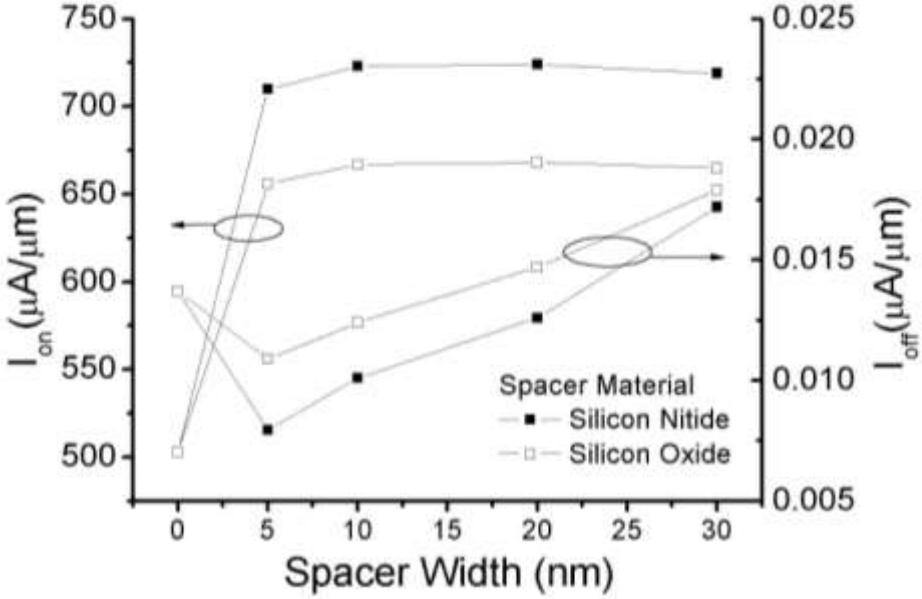
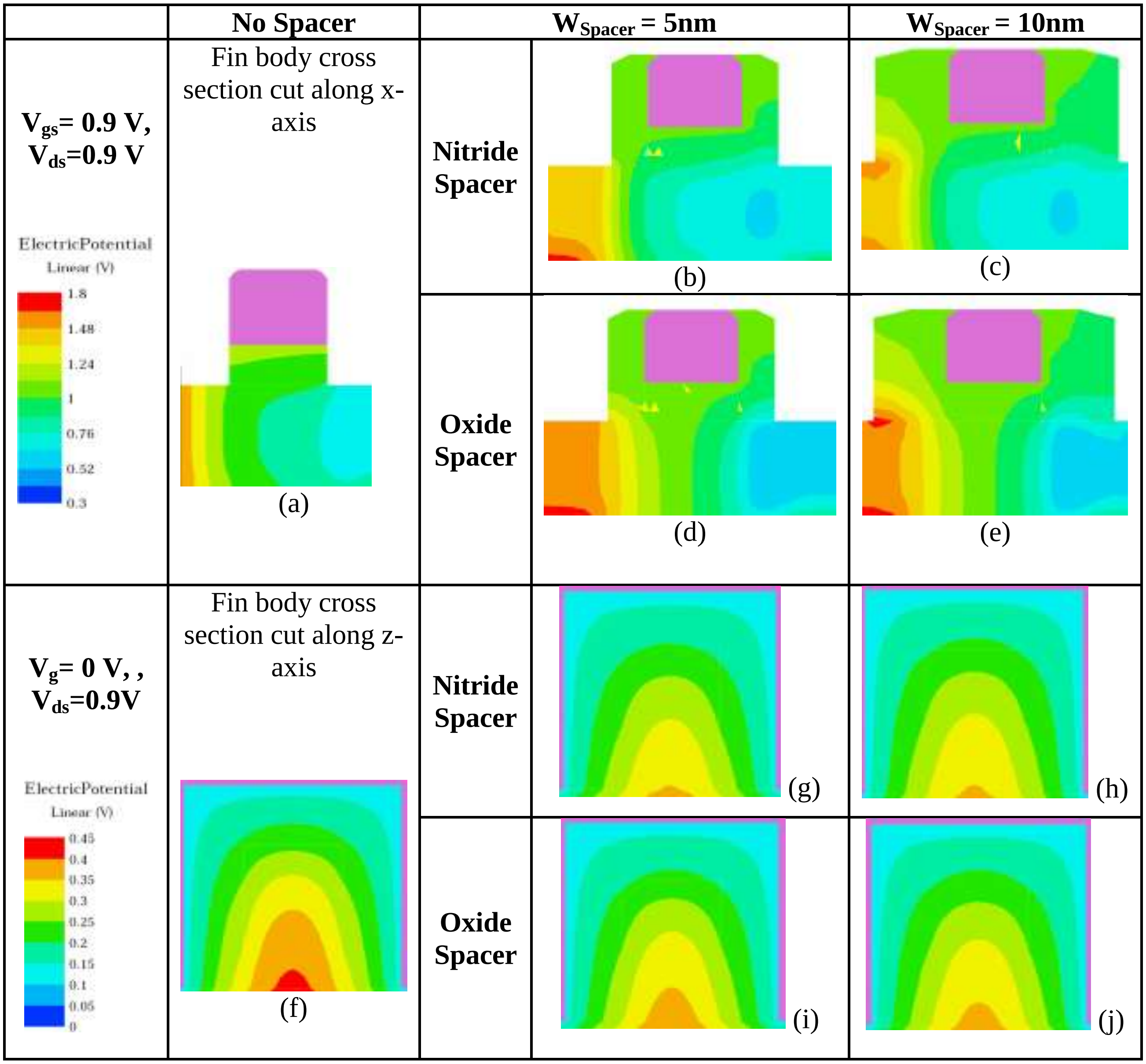


**Fig. 2-9**: On- and Off- currents for devices with different spacer widths and materials. Both devices have volume gate electrode. Fin width is 9 nm and gate length is 13 nm.

The data in **Fig. 2-7** and **2-8** suggest that for FinFET device with under lapped gate structure, the electrostatic can be strongly influenced by the gate and the spacer geometry in the proximity of the channel through gate-to-source/drain capacitance $C_{gs}$. Further, both the spacer width and the permittivity of the spacer material should have strong impact on the drain current. **Fig. 2-9** plots the $I_{ON}$ and $I_{OFF}$ for device with various spacer widths and two spacer materials, silicon oxide (ε=3.9) and silicon nitride (ε=7.9). It can be seen that the $I_{ON}$ is relatively constant for all spacer widths larger than the gate under lap width (~5 nm in this case) although that of the device without spacer (i.e. the 0 nm point) is much lower. This is attributed to both the nature of under-lap gate geometry and the fringing field through spacer. Since the lowly doped channel in the under lap region is not directly under the control of the gate field, the electrostatic potential in this region is especially sensitive to fringing electric field from gate through spacer in its proximity.



| | No Spacer | $W_{Spacer}$ = 5nm | | $W_{Spacer}$ = 10nm |
|---|---|---|---|---|
| $V_{gs}$= 0.9 V, $V_{ds}$=0.9 V | Fin body cross section cut along x-axis<br>(a) | Nitride Spacer | (b) | (c) |
| | | Oxide Spacer | (d) | (e) |
| $V_g$= 0 V, , $V_{ds}$=0.9V | Fin body cross section cut along z-axis<br>(f) | Nitride Spacer | (g) | (h) |
| | | Oxide Spacer | (i) | (j) |

**Fig. 2-10**: Electrostatic potential plot for (a)-(e): cross section along x-axis through mid-fin at bias Vds=0.9 V, Vgs=0.9 V. (f)-(j): cross section perpendicular to x-axis at center of channel at bias Vds=0.9 V, Vgs=0 V. All devices are identical except for the width of spacers. Fin body under gate has higher potential for devices without spacer due to reduced fringing capacitance from gate sidewall to S/D region and under lapped channel region.

A detailed analysis of the electrostatic potential distribution in the channel region can be helpful in understanding this. In the potential plots (a)-(e) in **Fig. 2-10**, the channel potential is plotted for devices with nitride spacer at bias $V_{gs}$=$V_{ds}$=$V_{dd}$. For the device without spacer shown in plot (a), the channel shows higher potential and smaller potential gradient; while for the device with spacer in plot (b), the channel potential is lower with larger gradient as a result of additional fringing electric field from gate through the spacer terminating on the fin body which is lightly doped near the edge of gate due to the underlap structure. Plot (c) shows that when the spacer is much wider, even extending into the highly doped S/D extension region, $I_{ON}$ remains almost constant because the channel potential is hardly affected. The same trend is observed in plots (d) and (e) which consider oxide as the spacer material. Further, oxide spacers yield lower $I_{ON}$ because the fringing field is less strong in this case due to lower permittivity of $SiO_2$. However, the $I_{OFF}$ is lowest when spacer width is approximately the same as gate under lap width. As shown in the cross sectional potential plots (f)-(j) in **Fig. 2-10** plotted for the off-state at bias $V_{gs}$=0 and $V_{ds}$=$V_{dd}$, it can be seen that the fin body potential is noticeably higher in the plot (f) which considers the device without the spacer, indicating lowering of source-channel barrier induced by the high drain bias leading to higher $I_{OFF}$. On the other hand, if the spacer is much wider than the length of under lapped channel as in plots (i) and (j), the potential at the Si surface above BOX starts increasing, induced by fringing field and the potential increases in the BOX and the substrate [27]. Therefore the $I_{OFF}$ increases due to back channel conduction from the bottom of fin body. The plots in **Fig. 2-10** also explain why devices with nitride spacer have less $I_{OFF}$ compared with those with oxide spacer of the same spacer width. As a result of higher permittivity of nitride, the fringing capacitances are higher in devices with nitride spacer even though the device geometry is the same. Body potential is affected less by fringing field for the oxide spacer cases and thus higher back channel potential and off state current. Based on **Fig. 2-7** to **-9**, one can establish that a thinner spacer of higher permittivity is beneficial to FinFET device with under-lapping gate for higher $I_{ON}$ and lower $I_{OFF}$.

Another component of parasitic in FinFET device is the series resistance from the highly doped fin body (or S/D extension region which is defined by the distance from raised source/drain contact region to the source/drain-to-channel junction, as shown in **Fig. 2-3**). Since a thinner spacer is preferred for devices with under lapping gate, technically the S/D extension region can be made shorter in order to reduce the series resistance. Our simulation shows that this is an effective design strategy for it benefits the current drive as well as the switching speed. As shown in **Fig. 2-11**, with increasing S/D extension length, the $C_{gg}$ increases until it saturates. Meanwhile the current drive is reduced due to higher series resistance. As a net effect of the two, the switching delay can be reduced by ~20% if S/D extension length is reduced from 40 nm to 20 nm.

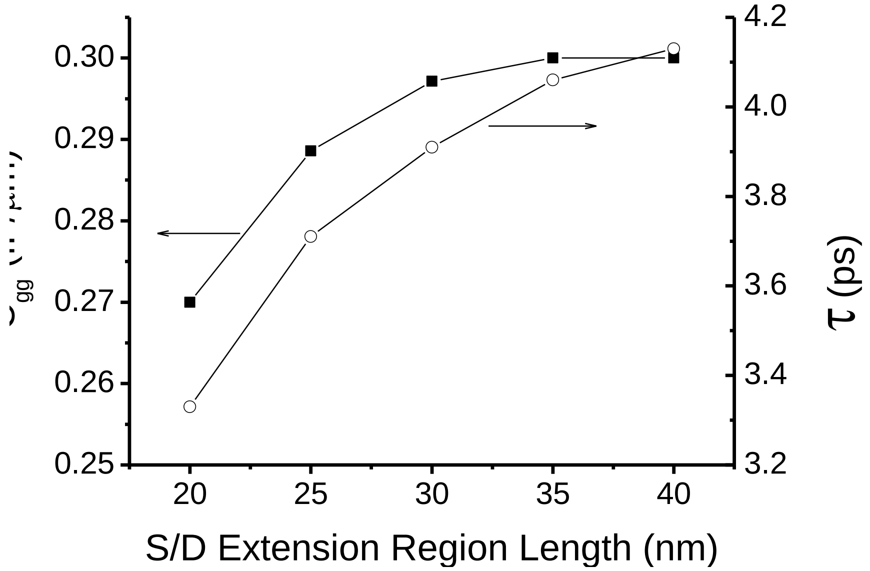


**Fig. 2-11**: Total gate capacitance and intrinsic delay plotted as a function of S/D extension region length. Shorter extension length is favorable for both enhanced current drive and reduced gate load capacitance. Bias condition: $V_{ds}$=$V_{gs}$=0.9 V. All devices have uniform spacer width of 10 nm.

## *2.4 Effects of High-k Gate Dielectric Material and Dielectric Thickness Scaling*

As the oxide gate dielectric has already reached its scaling limit, an alternative high-κ gate material is needed for the sub-32nm generation devices. The effect of replacing $SiO_2$ with high-κ gate dielectric material is studied in this section. Though some of the EOT studied here may not be feasible for $SiO_2$ due to its high leakage issue [28], the conclusion drawn for the device behavior on the basis of 3D simulations with high- and low-κ materials would still be representative.

Cheng *et al* [29] and Kumar *et al* [30] have discussed the effect of high-κ gate dielectric material in planar short channel SOI devices with spacer and overlapping gate based on 2D simulations. It has been argued that the when κ value of the gate dielectric material increases and device channel length continues to scale down, the gate dielectric thickness becomes comparable to the device channel length, resulting in increased fringing fields from the gate to the S/D region resulting in more severe $V_t$ roll-off, and degraded SCEs. However, both [29] and [30] only considered the fringing component from the bottom of the gate to S/D [i.e., component 2 in **Fig. 2-3**]. In the case of sub-20 nm channel length FinFET, three-dimensional electric field and electrostatics are important for $C_{gs}$ calculations as we have shown in the previous section. This is because the part of "under lapped channel region" is not directly under the control of gate electric field, and therefore it is extremely sensitive to all fringing field depicted in **Fig. 2-3**. By using more realistic 3D simulation, we are able to draw a more complete comparison of the capacitance, current, and intrinsic delay in devices with respect to the scaling of high-κ device.

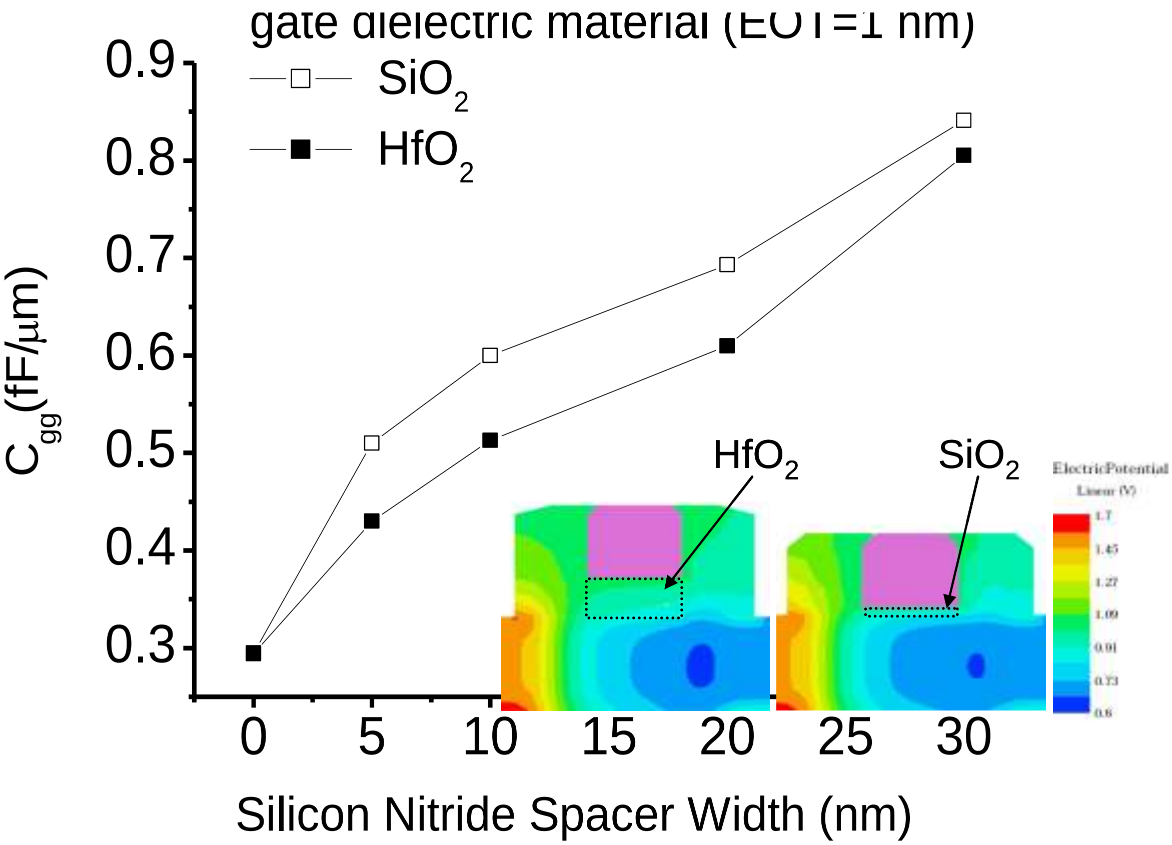


**Fig. 2-12**: Total effective gate capacitance with respect to silicon nitride spacer width for devices with $SiO_2$ and $HfO_2$ (ε=21) gate dielectric material of the same EOT (1nm). All devices have fin width 9 nm, gate length 13 nm, fin body thickness 13nm. Bias condition: $V_d$=$V_g$=0.9V and $V_s$=0V. Inset: electric potential plot in channel comparing devices with different dielectric of same EOT both device has same spacer with of 10nm.

As shown in **Fig. 2-12**, in case of FinFET with gate under-lap, the device with high-κ gate dielectric layer actually shows ~20% reduction in total gate capacitance compared with that of oxide gate dielectric for all spacer widths. This loss of total gate capacitance is mainly attributed to the spatial redistribution of electric field when a dielectric of much larger physical thickness is used in place of thin low-κ $SiO_2$. As a result of the this redistribution [29] the vertical electric field strength in dielectric layer and channel are reduced, since less field lines originating from gate electrode terminate in channel of device with high-κ dielectric compared with that with oxide dielectric layer. Consequently, the electrostatic potential distribution, as plotted in inset of **Fig. 2-12**, is significantly affected by the physical thickness of the gate dielectric although the EOT is maintained. For the device with oxide gate dielectric, the potential in Si body under the gate has lower potential with a high potential gradient. From this figure, it is not surprising that the 3D simulation of FinFET with under-lapped gate shows that by replacing oxide with a high-κ material of the same EOT leads to degraded $I_{ON}$ and increased $I_{OFF}$.

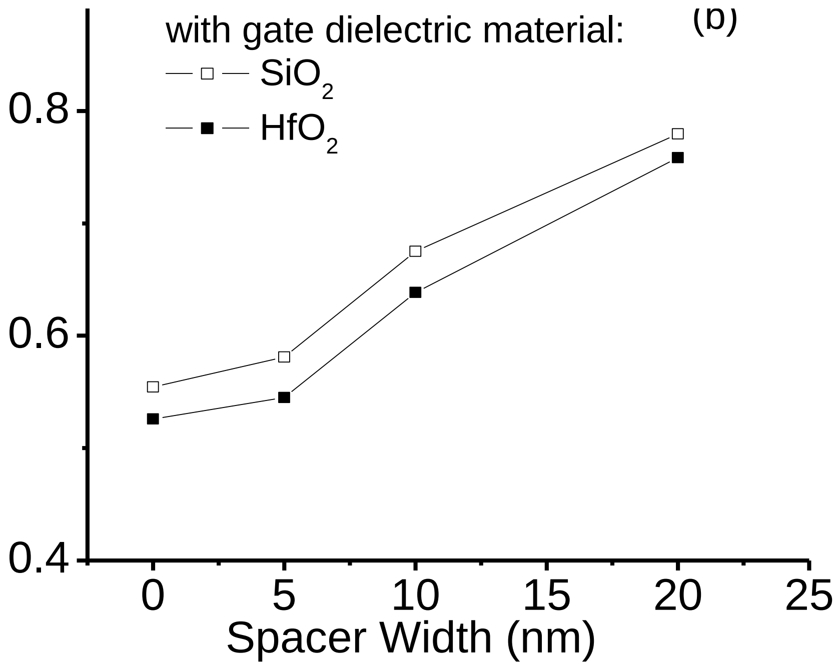


**Fig. 2-13**: Intrisic gate delay with respect to silicon nitride spacer width for devices with $SiO_2$ and $HfO_2$ ($\varepsilon$=21) gate dielectric material of the same EOT (1nm). All devices have fin width 9 nm, gate length 13 nm, fin body thickness 13 nm. Bias condition: $V_d$=$V_g$=0.9 V and $V_s$=0 V

**Fig. 2-13** examines the intrinsic gate delay of devices with $SiO_2$ and $HfO_2$ gate dielectrics and silicon nitride spacer with varying spacer width as a net result of fringing capacitance and the current drive. In terms of the switching speed, both operate slower for wider spacer due to stronger gate fringing field and the devices with $HfO_2$ gate dielectric are faster by ~10% for the same spacer width.

In conventional device with $SiO_2$ gate dielectric, reducing dielectric thickness causes improvement in many aspects of device performance such as the current drive and subthreshold leakage current. Scaling down the high-κ dielectric thickness might be an option for enhancing the device performance further in FinFET. In rest of this section we investigate the effectiveness of this option using $HfO_2$ as an example.

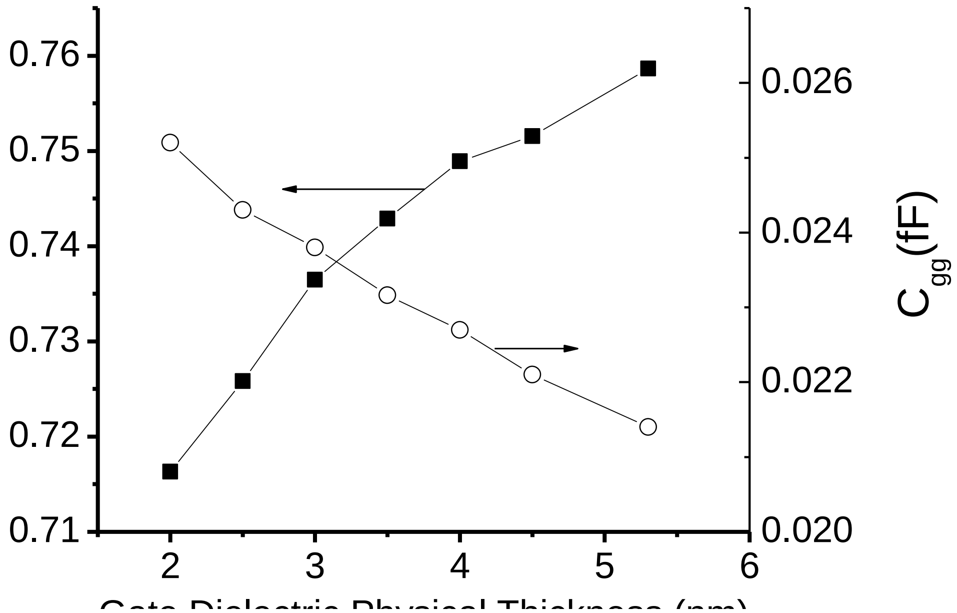


**Fig. 2-14**: Variation of intrinsic delay and $I_{ON}$ normalized by dielectric capacitance with respect to high- κ dielectric ($HfO_2$, ε=21) thickness.

As the dielectric thickness is scaled down and its capacitance value $C_{diele}$ increases, it is expected that the device short channel performance, off state S/D leakage current, and current drive will improve with decreasing dielectric thickness. However, from the circuit performance point of view, the incentive of reducing high-κ thickness needs to be examined carefully. The simulations presented here indicate that the down-scaling of gate dielectric thickness by 2.5 times results in only ~26% enhancement in the drive current. The carrier distributions in the channel region suggest that this less than expected current enhancement is due to the enhanced quantum effects when very thin dielectric layer is used as gate insulator. This is validated by the calculation of EOT based on the MLDA model adopted in the simulation taking into account the shifting of locations of peak carrier concentration. The EOT is increased by ~ 0.27 nm when a 4.5 nm $HfO_2$ dielectric layer is used (total effective EOT ~1.11nm). The EOT increases by as much as ~0.56 nm for a 2 nm $HfO_2$ layer (total effective EOT ~0.93nm). The current improvement is therefore limited. Further, Lombardi's model indicates mobility degrades near the source end due to enhanced vertical (normal to channel) field in case of very thin dielectric layer. On the other hand, $C_{gg}$ also changes moderately on the same account. It increases by ~18% (as seen in **Fig. 2-14**) after reducing the gate dielectric thickness by 2.5 times. The crowding of field lines that leads to high fringing capacitance also becomes less significant as the dielectric thickness becomes smaller compared to the gate length [30], leading to further reduction in $C_{gg}$ value. As a result, the net effect on the device intrinsic gate delay remains favorable: ~6% decrease as shown in **Fig. 2-14**. Thus, the down-scaling of the high-κ thickness for FinFET can be an approach for device performance enhancement if the device structure are carefully designed and evaluated as the implementation of high-κ dielectric remains indispensable for 32nm technology node due to gate leakage current issue.

### *2.5 Scaling of Fin Width, Gate Electrode Thickness, and Pitch in Multi-fin Devices*

FinFET devices typically have smaller width compared with their planar counter part due to aspect ratio constraint of the vertical fins. It will be desirable in circuit applications to connect multiple fins in order to enhance current drive. In this section, we discuss the effect of the fin width ($W_{fin}$) and fin pitch on the device performance. For a typical gate material (poly, FUSI and metal gate with poly-Si capping layer [31]), the gate electrode material thickness will be larger than fin height for small fin geometries. Therefore the space between adjacent fins will be filled up by gate electrode as long as the pitch is not much larger than twice of the thickness of the gate electrode material [32]. As shown in **Fig. 2-1**, for a multiple-fin device, this width between adjacent fins is the actual gate electrode thickness $H_{gate}$ for the two parallel gates residing at the sidewalls of the fin body. Thus, for a design of a certain fin width, the only top gate electrode thickness is dependent on the gate material thickness while the side gate electrode thickness is often not a choice once the fin pitch is decided.

Vertical conducting channels in close proximity to each other will be another source of fringing capacitance for multiple-fin devices. The geometry dependent parasitic capacitance for multi-fin double gate devices was analyzed by Wu & Chan [33] using a pseudo-3D conformal mapping technique. We wish to mention here that this pseudo 3D approach yields a gate capacitance of ~0.411 fF/µm while the full 3D simulations give a $C_{gg}$ value of 0.57 fF/µm. Thus simplifying 3D electric fields into several 2D fields, may have only limited accuracy. And, this inaccuracy increases with increasing fin pitch bringing out the importance of full 3D analysis.

We examine the effect of pitch on fins of various widths from 4 nm to 18nm while fixing the fin height at 13 nm. $HfO_2$ with EOT of 1 nm is used for gate dielectric and the nitride spacers have a width of 20 nm. The gate electrode thickness is 100 nm. **Fig. 2-15** shows how the total gate capacitance varies with fin width and pitch. We can deduce two conclusions from this plot. First, the total capacitance peaks when $W_{fin}$ is approximately half of the fin height but the dependence is not very strong. Second, fin pitch/gate electrode thickness has a strong effect on the capacitances.

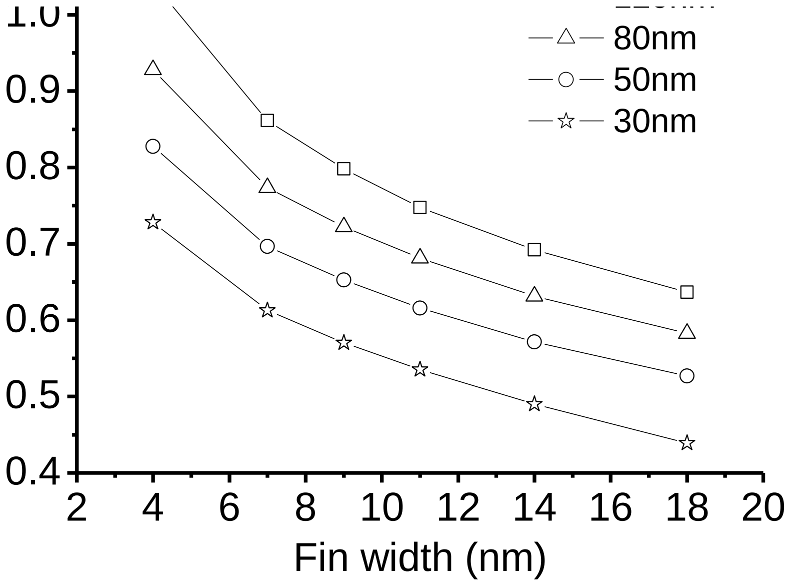


**Fig. 2-15**: Intrinsic delay against fin widths for multi-fin devices of various pitches. All devices have gate length 13nm, gate and spacer height 20nm, and the fin body thickness 13 nm. Bias condition: $V_d$=$V_g$=0.9V and $V_s$=0V.

For a given fin width, increasing the pitch by 4 times results in 50% increase in effective gate capacitance. This is mainly due to increase in $C_{gs}$ and gate-to-substrate capacitance $C_{gb}$ due to increased gate electrode area, enhancement in $C_{gs}$ contributes 72% and increased $C_{gb}$ contributes 28% to the increase in $C_{\mathrm{gg}}$ in this case. The same trend was observed in [33] by conformal mapping of 2D simulation. However, since the model in [33] was based on double-gate device with overlapping gate geometry, the percentage increase in capacitance is much less with increasing pitch. It is also predictable that with this increase in capacitance, the device switching speed will be slowed down. As shown in **Fig. 2-16**, increasing the fin pitch indeed results in longer intrinsic delay for all fin widths due to the higher fringing capacitance from larger $H_{\mathrm{gate}}$. A wide fin with small pitch has the shortest gate delay. However, this is at the expense of the $I_{\mathrm{OFF}}$ which is significantly higher in wider fins. A well-designed device should have a good compromise between the speed and the standby power.

With the development of novel gate stack technology, very thin and highly conductive metal gate electrode (2nm~10nm) is now made possible with atomic-layer deposition [34]. The direct effect of such a reduction of gate electrode thickness is the reduction in fringing field from sidewall of gate electrode to S/D as shown in the inset of **Fig. 2-16**. Simulation shows that this benefits the device performance by reducing parasitic capacitance without affecting the current drive. As shown in **Fig. 2-16**, for higher fin pitch that is larger than twice of the fin width, thin gate electrode yields about 10% improvement in speed.

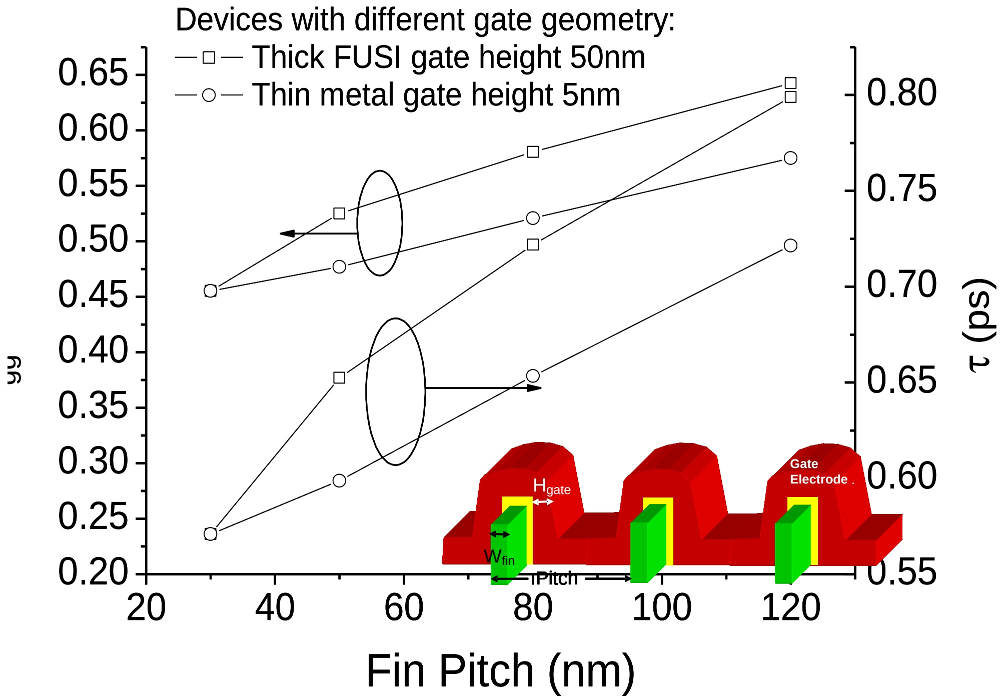


**Fig. 2-16**: Comparing intrinsic delay and total effective gate capacitance with respect to fin pitch for devices with thick FUSI gate electrode and thin metal electrode. Identical gate workfunction of 4.6 eV are assumed for all devices. All devices have gate length 13 nm, and the fin body thickness 13 nm. Inset: a multi-fin device with very thin gate electrode that is much less than half pitch. Spacers, BOX and substrate Si layers are not shown for clarity

From **Fig. 2-11, 14, 15** and **16**, it is striking that geometry factors such as fin width, S/D extension region length, and spacer width as well as gate dielectric thickness all have a significant impact on the performance, especially switching speed of FinFET devices.

This further highlights the importance of fringing field in very small FinFET devices. As FinFET devices are projected to be helpful in continuing the scaling trends for MOSFET devices, structural parameters, such as pitch and gate electrode thickness that are non-critical in conventional devices become more important in FinFET design and optimization. These parameters, together with the fringing field effects related to them, should be considered carefully in the design and modeling of the FinFET devices.

## *2.6 Implications to Nanowire and Other Nanoscale MOSFET Devices*

The key concept behind ultrathin-body transistor is improved SCE. As we have shown in the previous sections, in order to achieve good SCE, the body thickness needs to be considerably thinner than the gate length. A family of thin-body devices evolves from FinFETs, including tri-gate FETs, Pi-gate FETs, Omega-gate FETs and Nanowire FETs in the order of increasing gate control [35] (as shown in the schematic in **Fig. 2-17**). As the degree of gate control increases, the electrostatics gets improved in multiple-gate structures as the gate influences the channel potential from more directions. Apparently, the gate-all-around (GAA) structure is the most resistant to SCE among all the emerging device structures for a given body thickness.

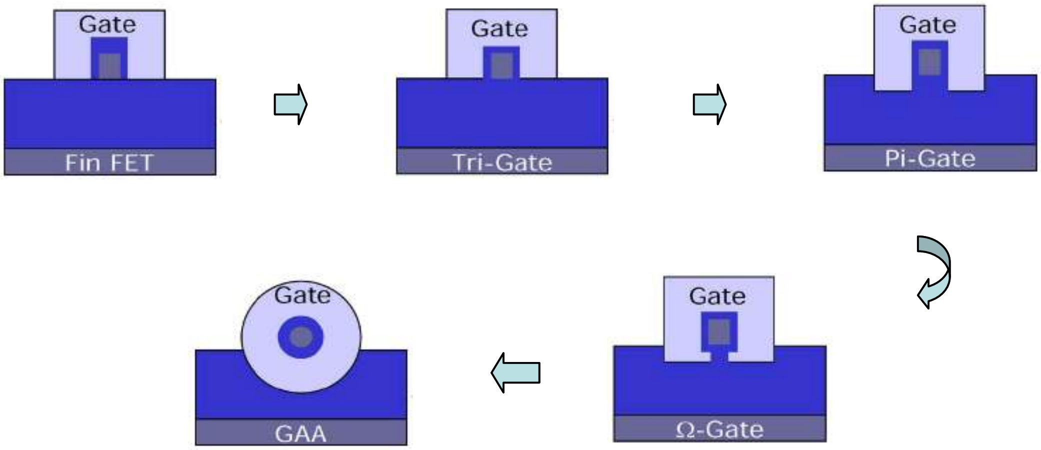


**Fig. 2-17**: Progression from FinFET to GAA NWFET [35]

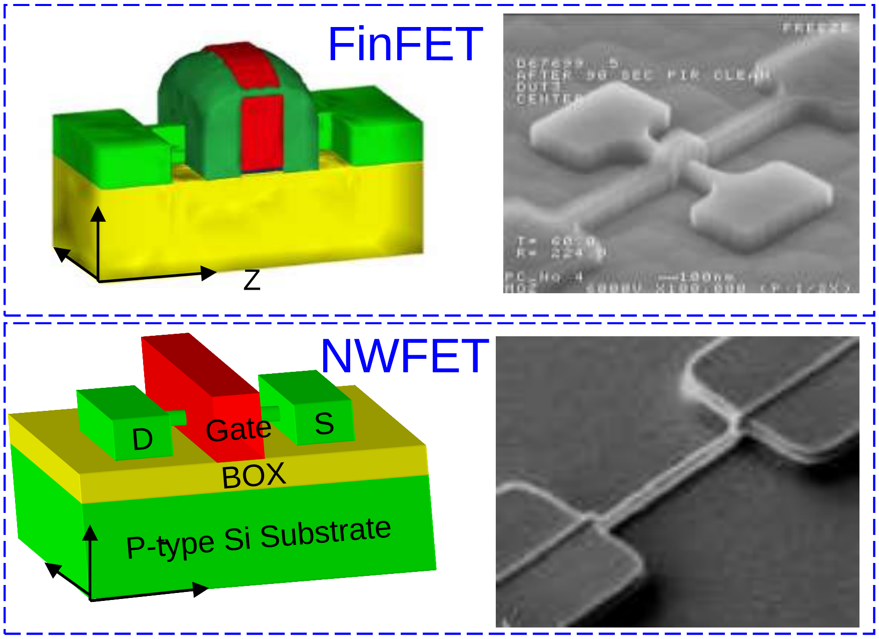


**Fig. 2-18**: Comparing the structural difference between FinFET and NWFET

Moving on from FinFET to NWFET seems to be a natural step of device evolution with simulation results [36] indicating that in cylindrical GAA architecture, the gate length can be scaled to 5nm with the corresponding scaling of the intrinsic NW channel diameter. However, in terms of device structures, there is more similarity than difference between a FinFET and NWFET. As shown in the device model and SEM image in **Fig. 2-18**, NWFET can be modeled using an adapted FinFET model with the channel suspending and surrounded by gate electrode instead of residing on the BOX layer. If we examine the two device structures closely, it is apparent that the three major fringing capacitance components from gate to S/D in FinFET depicted in **Fig. 2-3** remains prevalent in NWFET as shown in **Fig. 2-19**. Therefore, it can be seen that fringing field would also play a significant part in NWFET devices. 3D device structural parameters analyzed in the previous sections, such as spacer width and material, gate electrode thickness as well as fin pitch should be applicable for under lapped gate devices through fringing electric field.

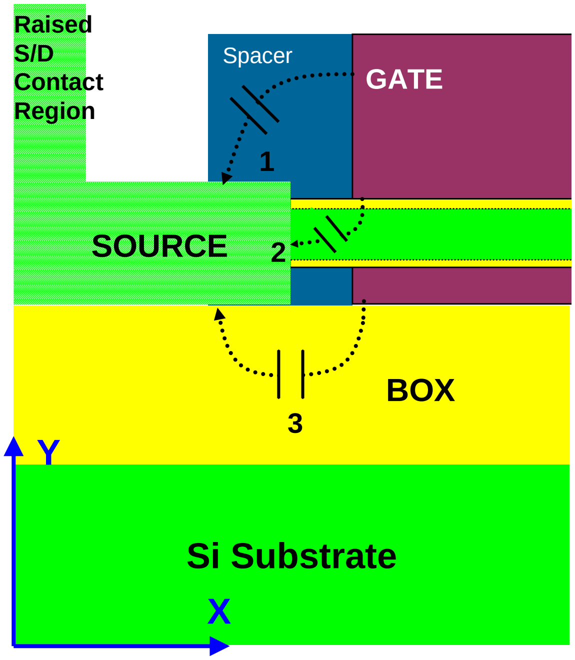


**Fig. 2-19**: A cross-section of the NWFET device along the centre of channel. Three major components of the gate-to-source capacitance $C_{gs}$ are illustrated

These effects in NWFET have been studied in a separate student project [37] and it has shown that the capacitance characteristics and trade-offs of device geometries, can be analyzed and optimized in a very similar fashion for NWFETs.

## 2.7 Chapter Summary and Conclusion

For the first time, a detailed analysis of device performance with consideration of 3D fringing capacitance is carried out for the FinFET devices of 32nm technology node. It is found that the 3D device structural parameters such as the spacer width and material, gate electrode thickness as well as fin pitch have significant effects for under lapped gate devices through fringing electric field. The performance impact of geometry factors was carefully analyzed by comparing channel electrostatic as well as the fringing capacitance resulting from geometry changes. Scaling down of the high-κ dielectric thickness is found favorable for device performance concerning SCE suppression and switching speed despite the relatively moderate enhancement in the drive current. The adoption of relatively thinner spacers of higher permittivity material is found to be beneficial to device performance due to the suppression of SCE. Very thin and highly conductive metal gate might be needed to reduce the gate-to-drain and gate-to-substrate fringing capacitances for faster switching. Further, large fin pitch/gate electrode thickness has an adverse effect on the performance of multi-fin devices. Thus multi-fin device with a higher fin density is favored for shorter intrinsic delay. Recognizing the important role fringing and parasitic capacitance plays in multi-gate devices with small geometry, full 3D simulation is not only important but virtually essential for evaluation of multi-gate FinFET and also GAA device scaling and design. The subsequent chapters in this thesis will explore the measuring and modeling aspects of the femto-farad scale capacitances in GAA devices.

## *Reference for Chapter 2*

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

[2] E. J. Nowak, I. Aller, T. Ludwig, K. Kim, R. V. Joshi, C.-T. Chuang, K. Bernstein and R. Puri, Turning silicon on its edge: Overcoming silicon scaling barriers with double-gate and FinFET technology, IEEE Circuits and Device Magazine, Vol. 20, no. 1, pp. 20-31, 2004.

[3] B. Yu et al. FinFET scaling to 10nm gate length, IEDM Tech. Dig. 421-4, 2002.

[4] Y. Liu, K. Ishii, T. Tsutsumi, M. Masahara and E. Suzuki, 2003. Ideal rectangular cross-section Si-Fin channel double gate MOSFET fabricated using orientation-dependent wet etching, IEEE Electron Device Lett. Vol. 24, pp. 484-486.

[5] V. Subramanian et al. Planar bulk MOSFETs versus FinFETs: An analog/RF perspective, IEEE Trans. Electron Devices, vol. 53, no. 12, pp. 3071-3079, 2006.

[6] Kazuhiko Endo et al. Fabrication of FinFETs by damage-free neutral-beam etching technology, IEEE Trans. Electron Devices, vol. 53, no. 8, pp. 1826-1833, 2006.

[7] F. Gamiz, M. V. Fischetti, Monte Carlo simulation of double-gate silicon-on-insulator inversion layers: The role of volume inversion. J. Appl. Phys. Vol. 89, no. 10, pp. 5478-5487, 2001.

[8] A. Khakifirooz and D. A. Antoniadis, On the mobility in ultrathin SOI and GOI, IEEE Electron Device Lett. Vol. 25, no. 2, pp. 80-82, 2004.

[9] M. M. Chowdhury and J. G. Fossum, Physical insights on electron mobility in contemporary FinFETs, IEEE Electron Device Lett., vol. 27, no. 6, pp. 482-485, 2006.

[10] P. Yang and P. K. Chatterjee, SPICE modeling for small geometry MOSFET circuits, IEEE Trans. Conputer-Aided Design Integr. Circuits Syst., vol. CAD-1, 1982.

[11] V. Trivedi, J. G. Fossum, M. M. Chowdhury, "Nanoscale FinFETs with gate-source/Drain underlap", *IEEE Trans Electron Devices*, vol. 52, pp. 56-62, 2005.

[12] Bansal, C. P. Bipul, R. Kaushik, Modeling and Optimization of fringing capacitance of nanoscale DGMOS device, IEEE Trans. Electron Devices, vol. 52, no.2, pp. 256-262, 2005.

[13] Jong-Tae Park, J.-P. Colinge and C. H. Diaz, Pi-gate SOI MOSFET, IEEE Electron Device Lett., vol. 22, no. 8, pp. 405-406, 2001.

[14] Sung Dae Suk et al., High performance 5nm radius twin silicon nanowire MOSFET (TSNWFET): Fabrication on bulk Si wafer, characteristics, and reliability, IEDM Tech. Dig., pp. 717-720, 2005.

[15] J. Y. Song, W. Y. Choi, J. H. Park, J. D. Lee and B. G. Park, "Design optimization of gate-all-around (GAA) MOSFETs", *IEEE Trans. Nanotechnology*, vol. 5, no. 3, pp. 186-191, 2006.

[16] J. G. Fossum, M. M. Chowdhury, V. P. Trivedi, T.-J. King, Y.-K. Choi, J. An, and B. Yu, "Physical insights on design and modeling of nanoscale FinFETs", *IEDM Tech. Dig.*, 2003, pp. 679-682.

[17] K. Tanaka, K. Takeuchi and M. Hane, "Practical FinFET design considering GIDL for LSPT device", *IEDM Tech. Dig.*, 2005, pp. 980-983.

[18] J.W. Yang, P. M. Zeitzoff, and H. H. Tseng, “Highly manufacturable double-gate FinFET with gate-source/drain underlap”, *IEEE Trans. Electron Devices*, vol. 54, no.6, pp. 1464-1470, 2007.

[19] Beckx et al. “Implantation of high-κ and metal gate materials for the 45 nm node and beyond: gate patterning development”” *Microelectronics Reliability* vol. 45, issue 5-6, pp. 1007-1011, 2005.

[20] A. Dixit, K. G. Anil, R. Rooyackers, F. Leys, M. Kaiser, N. Collaert, K. De Mayer, M. Jurczak and S. Biesemans, “Minimization of specific contact resistance in multiple gate NFETs by selective epitaxial growth of Si in the HDD regions”, *Solid State Electronics*, vol. 50, pp. 587-593, 2006.

[21] N. Collaert et al. “Integration challenges for multi-gate devices”, *IEEE International Conference on Integr. Circuit and Tech*. 2005, pp. 187-194, 2005.

[22] N. Collaert et al., “Tall Triple-gate devices with TiN/$HfO_2$ gate stack”, *VLSI Tech. Dig. 2005,* pp. 108-109, 2005.

[23] Taurus Device™ and Taurus Process™ User Manual, Synopsys, 2007

[24] H. Zhao et al., “Simulation of multiple gate FinFET device gate capacitance and performance with gate length and pitch scaling”, *IEEE SISPAD Tech. Dig.*, pp. 252-255, 2006.

[25] H.-S. Wong, K.-W. Ang, L. Chan, K.-M. Hoe, C.-H. Tung, N. Balasubramaniam, D. Weeks, T. Landin, J. Spear, S. G. Thomas, G. Samudra and Y.-C. Yeo, “Source/Drain-extension-last process for incorporating in situ doped lattice-mismatched extension stressor for enhanced performance in SOI N-FET”, *ISDRS 2007,* pp. 1-2.

[26] A. T. Tilke, L. Pescini, M. Bauer, M. Stiftinger, R. Kokoschke, D. Shum, N. Chan, S. Kim, V. Hecht, and K. J. Han, “Highly scalable embedded flash memory with deep trench isolation and novel buried bitline integration for the 90-nm node and beyond”, *IEEE Trans. Electron Devices*, vol. 54, no. 7, pp. 1681-1688, 2007.

[27] T. Ernst and S. Cristoloveanu, Buried oxide fringing capacitance: a new physical model and its implication on SOI device scaling and architecture, IEEE SOI Conference Proc. 1999, pp. 38-39.

[28] Y. Taur, CMOS design near the limit of scaling, IBM J. Res. & Dev., vol. 46, no. 2/3, pp. 213-222, 2002.

[29] B. Cheng et al., The impact of high-gate dielectrics and metal gate electrodes on sub-100nm MOSFETs. IEEE Trans. Electron Devices, vol. 46, no. 7, pp. 1537-1544, 1999.

[30] M. J. Kumar, S. K. Gupta and V. Venkataraman, Compact modeling of the effects of parasitic internal fringing capacitance on the threshold voltage of high-κ gate-Dielectric nanoscale SOI MOSFET, IEEE Trans. Electron Devices, vol. 52, no. 4, pp. 706-711, 2004.

[31] H. J. Cho et al., The effects of TaN thickness and strained substrate on the performance and PBTI characteristics of poly-Si/TaN/HfSiON MOSFETs, IEEE IEDM Tech. Dig. 04-503, 2004.

[32] T. Ludwig, I. Aller, V. Gernhoefer, J. Keinert, A. Mueller, E. Nowak, R. V. Joshi, and S. Tomaschko, FinFET technology for future microprocessors, IEEE SOI Conference Tech. Dig., pp. 33-34, 2003.

[33] W. Wu and M. Chan, Analysis of geometry-dependent parasitics in multifin double-gate FinFETs, IEEE Trans. Electron Devices, vol. 54, no. 4, pp.692-698, 2007.

[34] S.-C. Song, Z. Zhang, C. Huffman, J. H. Sim, S. H. Bae, P. D. Kirsch, P. Majhi, R. Choi, N. Moumen and B. H. Lee, Highly manufacturable advanced gate-stack technology for sub-45-nm self-aligned gate-first CMOSFETs, IEEE Trans. Electron Devices, vol. 53, no. 5, pp. 979-989, 2006.

[35] N. Singh, K. D. Buddharaju, S. K. Manhas, A. Agarwal, S. C. Rustagi, G. Q. Lo, N. Balasubramanian, and D.-L. Kwong, "Si, SiGe nanowire devices by top-down technology and their applications", *IEEE Trans. Electron Devices*, vol. 55, no. 11, pp. 3107-3118, 2008.

[36] E. Gnani, S. Reggiani, M. Rudan, and G. Baccarani, "Design considerations and comparative investigation of ultra-thin SOI, double-gate and cylindrical nanowire FETs", in *Proc. IEEE ESSDERC*, 2006, pp.371-374.

[37] B. Chen, "Simulation of Nanowire MOSFET", Thesis Dissertation for Degree of Bachelor in electrical engineering, National University of Singapore, 2009.

# Chapter 3

# Charge Based Capacitance Measurement (CBCM) for Femto-Farad Scale Capacitance Measurement

# – Simulation and Analysis

## *3.1 Introduction*

### *3.1.1 Challenges in C-V characterization of nanoscale devices*

NWFET devices, among other ultra-thin body devices, are expected to play a critical role in the future semiconductor technology nodes due to its compatibility to standard CMOS technology and the superior performance reported in SOI ultra narrow NW devices [1]. However, the much speculated effects of reduced density of states in 1D structure [2], strain induced bandgap change [3], and phonon-limited mobility reduction [4] are intriguing and the need for experimental confirmation is imperative. Accurate measurement of ultra-small capacitance is crucial to successful device characterization and development of compact models for circuit applications of the next-generation nanoscale devices such as NWFETs and FinFETs. Based on our knowledge of the field and scaling properties of FinFET and NWFET devices from previous chapter, the challenges being faced are, but not limited to, as follows:

1) Due to ultra-thin body structure, the capacitance of a single transistor becomes extremely small, as seen in previous chapter, typically in the range of hundreds of atto-farad (aF) which is much smaller than the resolution of the most of measurement devices.

2) Due to the 3-D nature of these devices, the electric field in 3-D is much different from that in the conventional planar devices. It is therefore not possible to characterize the capacitance using a planar large-area capacitance as in conventional VLSI technology.

3) As the device body thickness and gate length scales down together, parasitic capacitances are playing a more significant role in device characteristics. Both the channel capacitance and the parasitic capacitance needs to be properly characterized for correct modeling of device operation.

When the major device dimensions are reduced below 32nm, variability in both dimensions and characteristics is inevitable. This variability also increases the complexity and difficulty of the characterization task and should itself be characterized.

### 3.1.2 Conventional measurement options and CBCM

Conventionally, quasistatic method, auto-balancing bridge method, and the s-parameter analysis method are the main options for measuring a capacitance.

A **quasistatic C-V measurement** instrument calculates the capacitance from measured current changes induced by a small change in voltage. The capacitance can then be derived by dividing the current with the voltage ramp rate as in $C=I/(\mathbf{d}V/\mathbf{d}t)$. While this method is simple to apply, the ramp rate of voltage is constrained by instrument as well as the response time of device to achieve equilibrium. It is clearly not suitable for femto-farad scale capacitance measurement as measuring the extremely low current with accuracy is in itself a problem. Also, this method can be severely affected by any leakage current in the case of measuring the gate capacitance of a MOSFET.

A **auto-balancing bridge** instrument measures both the voltage across and the current through the device under test (DUT). An oscillator signal is output through the high current terminal of the DUT and can be changed to vary the signal level applied to the DUT. An op-Amp forming a “null-loop” virtually grounds the low potential terminal of the DUT and pulls the DUT current through a “range resistor”. By detecting the voltage across the range resistor, the current flowing through DUT can be measured. Because the voltage across DUT depends on the relative impedance of DUT and the source resistor, which is typically 50 Ω, the voltage applied across DUT can be much lower than applied signal level for very low DUT capacitors. In this case the signal to noise ratio (S/N) deteriorates significantly.

Finally, there is also **s-parameter analysis method,** which calculates the reflection coefficient by measuring the ratio of the complex power of the incident signal to the reflected signal. The reflection coefficient varies greatly with load impedance. When the measured impedance is very close to the internal impedance (typically 50 Ω), the highest accuracy is obtained. The impedance measurement sensitivity degrades as the DUT impedance becomes much lower or higher than 50 Ω. For nanowire type of DUT which has very low capacitance, impedance mismatch is almost inevitable. Therefore s-parameter analysis is clearly not the choice method for their C-V characterization.

Moreover, the conventional CV measurement instruments face tremendous difficulty in measuring ultra-small capacitances intrinsic to NW (aF to fF level) over large background parasitic capacitance (pF level).

Until now, the assessment of capacitance in ultra-thin body nanoscale devices such as NWFETs relies mostly on simulation [5]-[7] or measurement of capacitance of a large number of devices connected in parallel [8]-[10]. Simulations need to be backed up by experiment data. On the other hand, although connecting a large number of devices in parallel brings the collective capacitance to a measureable level, it also eclipses the variability in nanoscale. Hence, both are not truly desirable for accurate characterization of capacitance in nanoscale devices.

Charge Based Capacitance Measurement (CBCM) technique was originally proposed by B. Hofflinger [11] for measuring the interconnect capacitance. It also emerges as a promising technique for measuring low capacitance of active devices [12] and may have the potential of accurate capacitance measurement of single channel nanoscale device. In CBCM technique, the voltage is applied to the DUT capacitor in a very simple and unambiguous way. By increasing the frequency of the input signals, the average current in CBCM can be increased, providing enhancement of S/N ratio. It also has a mechanism to effectively remove the background capacitance without involving complicated shielding. However, there are hardly any reports in the literature assessing the accuracy of this technique and the parameters influencing it. In this chapter, we investigate the mechanism and accuracy of the technique specifically for sub-Femto Farad nano-wire based device capacitance by carrying out extensive mixed device and circuit-mode simulations. The factors that influence the accuracy of the technique are identified.

## 3.2 Principle of CBCM and Setup

CBCM is based on the well known switch capacitor (SC) principle: in a SC circuit in which clocked switches charge and discharge a well defined capacitance forms a circuit element with a defined resistive behavior. Vice versa, if the resistive behavior can be measured, the capacitance value can be determined. **Fig. 3-1** shows a much idealized model of the test structure. The two switches are clocked alternately with constant frequency $f$=1/$T$, charging the capacitance of device under test (DUT) $C_{DUT}$ to voltage $V$ applied by the voltage source or discharging it to the ground. If the voltage $V$ changes with a rate that is less than the frequency $f$ of the clock signals, the load to voltage source has a resistive property which can be expressed as $R=T/C_{DUT}$ because with in any one time interval $T$, the charge passing through either switch in $C_{DUT}{\cdot}V$. Therefore, if $T$, $V$ and the time averaged current $I_0$ are know, $C_{DUT}$ can be determined by:

$$C_{DUT}{\cdot}V= T{\cdot}I_0 \qquad (3\text{-}1)$$

If $C_{DUT}$ is voltage dependent, a differentiation with respect to the voltage is necessary to obtain $C_{DUT}(V)$:

$$C_{DUT}=T{\cdot}dI_0/dV \qquad (3\text{-}2)$$

To perform this C(V) measurement, the voltage source has to be swept with sweep frequency smaller than that of the two clock signals $f$ for $S_1$ and $S_2$ and the time averaged current $I_0$ has to be differentiated for each sweep step.

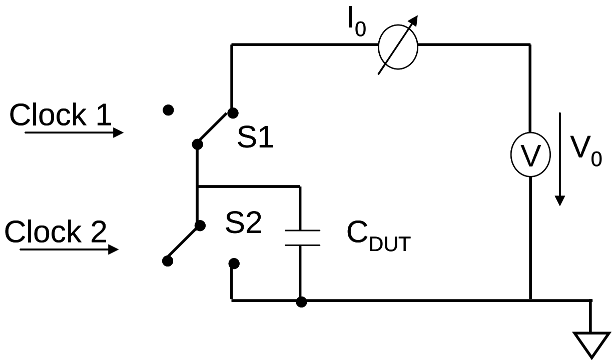


**Fig. 3-1:** Switch capacitance model for CBCM test

**Fig. 3-2 (a)** shows the CBCM implementation using NWFETs. Following the SC principle described, $S_1$ and $S_2$ can be implemented using a p-type and an n-type NWFET forming a pseudo inverter (shown as *N Driver* and *P Driver*).

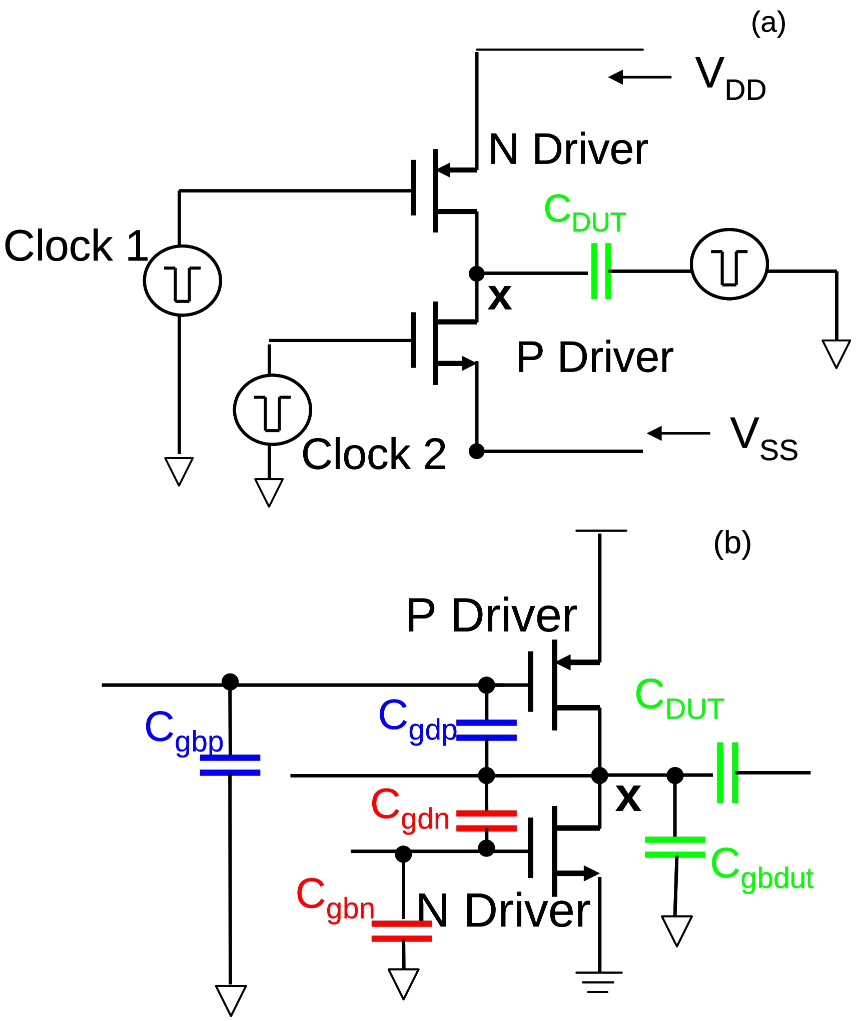


**Fig. 3-2** (a): CBCM implementation circuit with input and output terminals (b): equivalent circuit with parasitic capacitances for CBCM measurement

Since NWFETs are not ideal switches and may have finite parasitic capacitances, these parasitic capacitances need to be taken into account. As depicted in the equivalent circuit in **Fig. 3-2 (b)**, there are three nodes through which the charging/discharging take place: the two gate nodes of the drivers, and the node *X* that connects the drains of the drivers to $C_{DUT}$. Thus the charging and discharging currents through the switches are contributed by gate capacitance $C_{DUT}$ and parasitic capacitance $C_{par}$ . As shown in **Fig. 3-2 (b),** $C_{par}$ is the sum of the two capacitances in parallel: (1) $C_{gdp}$ and $C_{gdp}$ in series to each other and (2) $C_{gbp}$ and $C_{gbp}$ in series to each other. These four components of $C_{par}$ accounts for the gate-to-drain and gate-to-substrate capacitances in the driver devices.

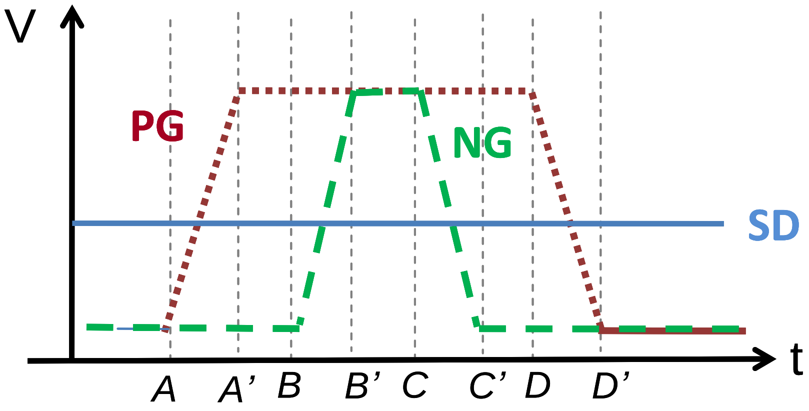


**Fig. 3-3**: The pulse shape for the input signals to the test key terminals during CBCM measurement. PG and NG are applied to the gates of the P and N-type driver respectively, a constant voltage SD is applied to the source and drain of DUT.

By switching the two drivers 'on' and 'off' alternately, the load capacitance comprising the $C_{DUT}$ and parasitic capacitance $C_{par}$ is charged to $V_{DD}$ when PMOS is turned 'on' and discharges to ground ($V_{SS}$) when NMOS is turned 'on'. Shown in **Fig. 3-3**, is the pulse shape of the input signals to the terminals of the test key: *PG* and *NG* are applied to the gates of the P and N-driver. They have the same frequency but different duty cycle. During the rising edge A-A', the P-driver is being turned off by *PG* while N-driver remains off and the voltage at Node X remains at *VDD*. During the rising edge B-B', the N-driver is being turned on by *NG* while P-driver remains off. This edge discharges the capacitance at node X to *0.*. During the subsequent falling edge C-C', the N-driver is being turned off by *NG* again while the voltage at node X remains at 0. The falling edge D-D' turns the P-driver back on, at the same time charging the capacitance at node X to back to *VDD*. Observe that at any instance, there is at most one driver turned on. Therefore, the current through *VDD* is not constant but varies time-dependently with the input pulse. The capacitance presented at node *X* is obtained from $Q_{VDD}$ which can be obtained by integrating the charging current $I_{V_{DD}}$ over the period of the pulses, or 1/*f* as in **Eqn. 3-3**

$$Q_{V_{DD}} = \int_0^{V_{DD}} (C_{DUT} + C_{par})dV = \int_0^{1/f} I_{V_{DD}} dt \tag{3-3}$$

Another charging current, $I'_{V_{DD}}$ which charges/discharges only the parasitic capacitances can then be obtained by repeating the measurement on a structure without the DUT [45]. Therefore the current $I'_{V_{DD}}$ only charges/discharges the parasitic capacitance, which can be extracted by:

$$C_{DUT} = \frac{d(Q_{V_{DD}} - Q'_{V_{DD}})}{dV_{DD}} = \frac{d(I_{V_{DD}} - I'_{V_{DD}})}{dV_{DD}} \frac{1}{f} \tag{3-4}$$

The most well known CBCM implementation for measurement of inter connect capacitance was proposed in [13], as shown in **Fig. 3-4 (a)**. One of the main problems with this CBCM setup using FET pseudo-inverter is the charge injection through $C_{gs}$ of P Driver during its switching-off process. The amount of charge injected to $V_{DD}$ depends on the rise/fall times of the pulses, magnitude of $C_{gd}$ and on the load to the pseudo-inverter. The charge injection issue has been addressed in two ways:

(i) by using the pseudo-inverters consisting of pass-gates instead of single transistors as shown in **Fig. 3-4 (b)** [14]. This design takes advantage of the compensation between positive and negative charges coming from the channels of P-FET and N-FET so that charge injection can be minimized.

(ii) by measuring the reference and DUT currents on the same pair of pseudo-inverters [15]. This setup measures $I'_{V_{DD}}$ by applying a third non-overlapping pulse to the source and drain of the DUT during the measurement, as an additional step. The added non-overlapping pulse is shown in **Fig. 3-4 (d)**. The third pulse signal rise/falls in between the two rising edge A-A’ and B-B’ as well as the two falling edge C-C’ and D-D’. It pre-discharges the source and drain of DUT to the same level as *VDD* before edge B-B’ discharge it to ground; it also pre-charges the source and drain of DUT before it is being charged by edge D-D’. In this setup, the DUT is proposed to remain un-charged during the measurement of the reference current $I'_{V_{DD}}$. Since the voltage between source and drain of DUT and node *X* is balanced before the drivers are turned on or off, the DUT is supposed to be electrically “invisible”, and not being charged and discharged.

To assess the efficacy of these techniques, TCAD simulations provide the best tool for evaluation as the instantaneous voltages on the nodes and currents in branches and electrostatics and carrier dynamics inside the device structure can be easily tapped.

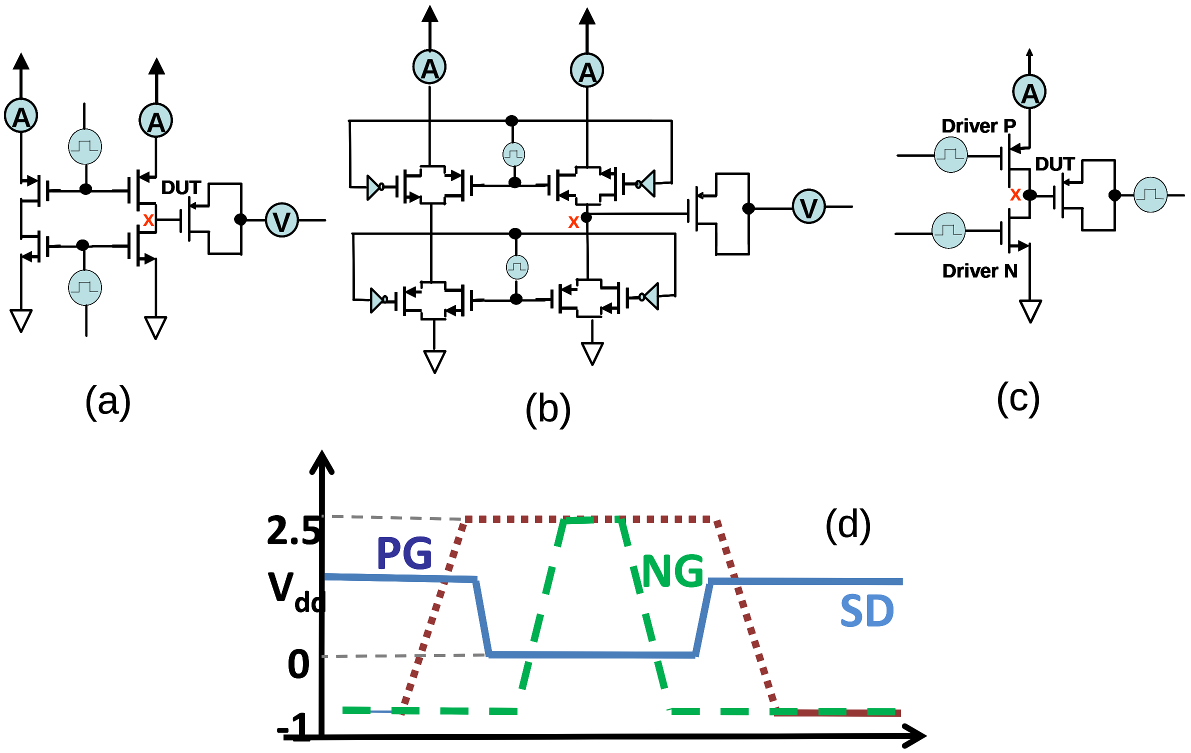


**Fig. 3-4**: CBCM test keys (a): proposed by Chen et al.[13] for interconnect capacitance (b): proposed by Vendrame et al [14]. Charge feed back in N- and P- transistors of the pass gates is expected to 'balances' out (c) proposed by Chang et al. [15]. It relaxes the demand on 'matching' as same drivers are used two times. (d): Input pulses PG – at P driver gate, NG – at N driver gate, SD – at S&D of DUT in setup (c).

### *3.3 CBCM Simulation and Efficacy Study of Three Setups*

To validate CBCM and study its efficacy, the CBCM setup described is implemented with mix-mode simulations in MEDICI. The DUT chosen is gate-all-around nanowire device that has a gate length of 0.25µm and cylindrical diameter of 10nm. 'WIDTH' parameter was used in MEDICI as the multiplicity parameter to simulate multiple "fingers" of DUT connected in parallel. The main advantage of using a physical DUT in place of a compact model is that physical DUT will automatically account for all the physics of charge movement including non-quasi-static effect, if any. On the other hand, N and PMOS drivers that essentially function as switches with attached junction and overlap capacitances can be well emulated by compact models. For all compact model drivers, we have gate length of 0.18μm and width of 0.5μm. This setup, thus, captures the nanoscale DUT physically, ensures proper convergence and economizes on computational resources. Repetition rate of 45 MHz, higher than that used in [14] and [15] was chosen in this study for two reasons: (i) more pronounced charge injection problem at higher frequency allows meaningful evaluation of the different methods; and (ii) higher frequency gives higher $I_{VDD}$, which would allow better accuracy during measurements Also, our simulations indicated insignificant effect of the frequency of input pulses on the extracted C-V characteristics of the DUT.

Shown in **Fig. 3-5** are the input pulses and the voltage tapped at node X, denoted by $V_{REF}$ and $V_{DUT}$ for reference branch and DUT branch respectively, when $V_{DD}$ applied is 0.8V. For setup (a) in **Fig. 3-4**, both $V_{REF}$ and $V_{DUT}$ shoot up significantly beyond $V_{DD}$ during the turning-off of the P driver, indicating impact of charge injection. As for the set up (b) in **Fig. 3-4**, the voltage overshoot of $V_{DUT}$ is lower compared with $V_{DUT}$ of (a) case (which is equivalent to the case when a constant voltage is applied to DUT in **Fig. 3-4 (a)**). Also worth noticing is the difference of the voltage levels for the two setups depicted in **Fig. 3-4 (a)** and **(c)** - although the difference is less in the latter, it has not been fully eliminated. More interesting is the case of setup (b) in **Fig. 3-4** which shows a little "undershoot" below $V_{DD}$ indicating a charge injection in the reverse direction, which is ascribed to the mismatch in the parasitic capacitors and threshold voltage of the two transistors comprising the pass gate.

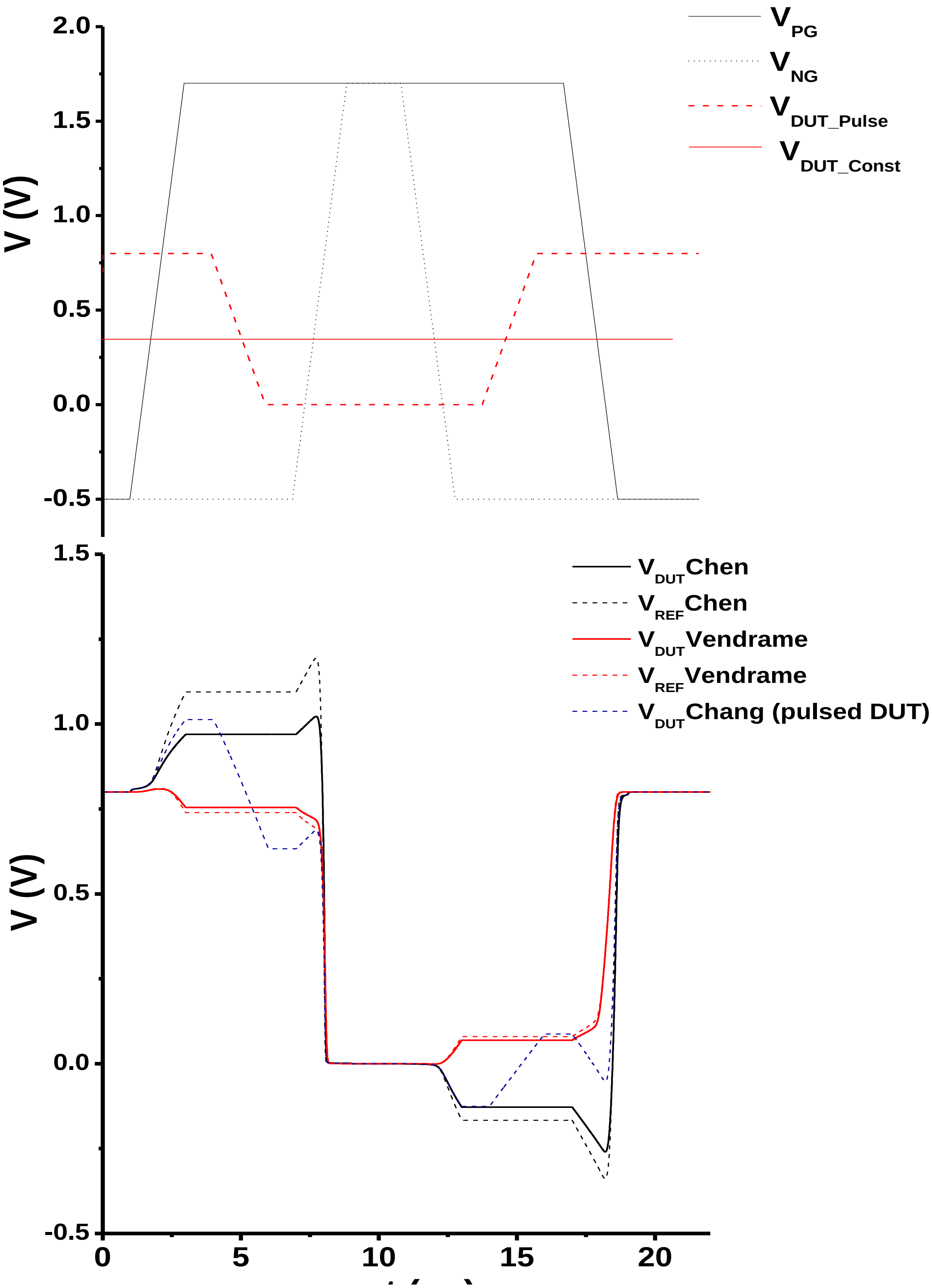


**Fig. 3-5**: Input and simulated output voltage wave forms for the three methods. For setups of Fig.3-4 (a) and (b), $V_{DUT}$ is at the gate of DUT and $V_{REF}$ is tapped at X in the reference branch. $V_{dd}$ is 0.8V. Setup in Fig. 3-4 (b) reduces overall amount of charge injected. Setup of Fig.3-4 (c) reduces difference in charge injection for two measurements when PMOS is switching off. The imbalance during the falling and rising edges of the pulse at S/D may not be significant as P-driver is off.

The bias-dependent capacitances of DUTs with various numbers of fingers derived using the three CBCM methods are shown in **Fig. 3-6** through **Fig. 3-8**. The simulation results in each case are compared with AC small signal simulation in MEDICI as the "actual" capacitance of the DUT. All three methods are accurate (within 3.5% of error) for 100 finger DUTs. However, when the DUT capacitances become comparable to or lower than the parasitic capacitance of the pseudo inverters, the effect of charge injection becomes more obvious and the derived capacitance becomes less accurate in all three cases.

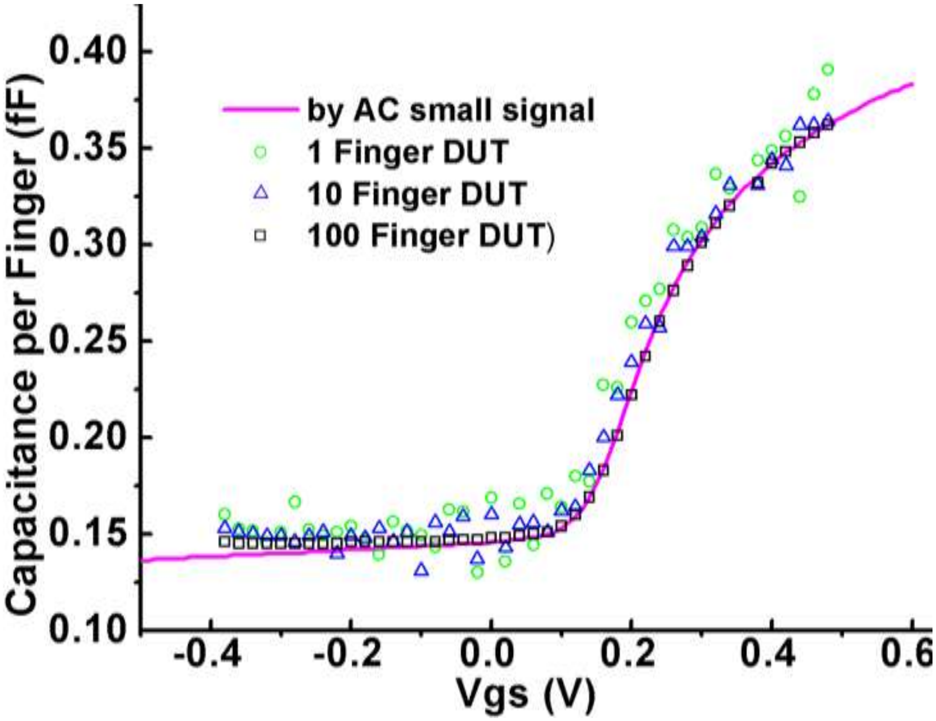


**Fig. 3-6**: Bias-dependent capacitance derived by CBCM setup of Fig 3-4 (a). DUTs with 1, 10, and 100 parallel fingers are considered and compared with the exact capacitance (small signal AC analysis in MEDICI). The error is very large for one finger DUT (max capacitance ~0.45fF).

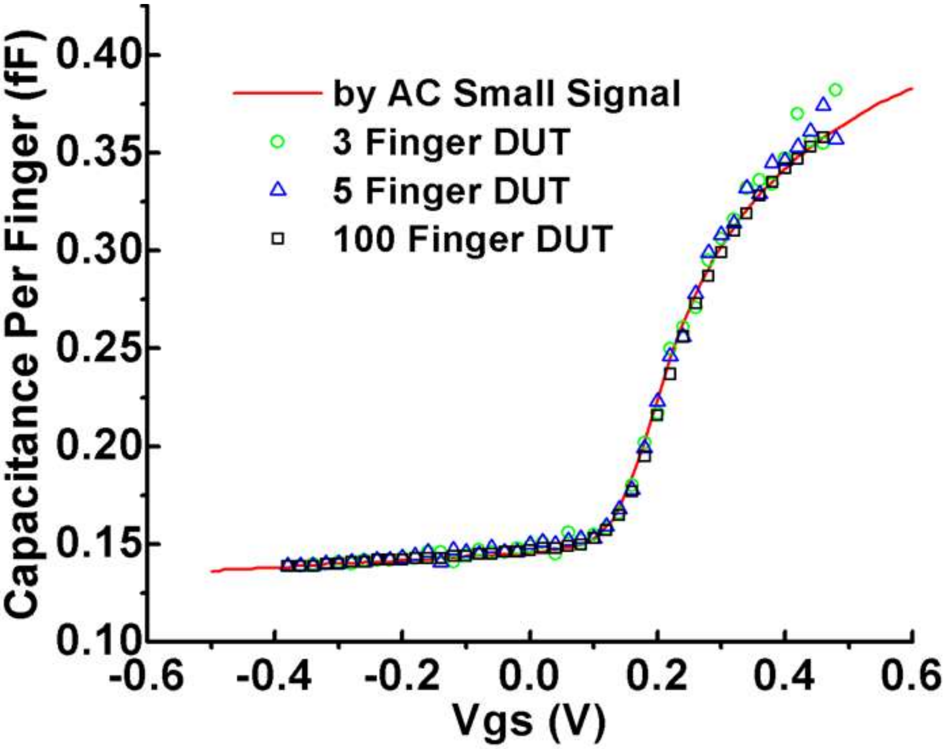


**Fig. 3-7**: Bias-dependent capacitance derived by the CBCM setup in Fig. 3-4 (b) with 5, 10, and 100 parallel DUT fingers. Solid line is the exact capacitance calculated by AC small signal method in MEDICI. This method is accurate even at 3 finger DUT (Max cap=1.35fF).

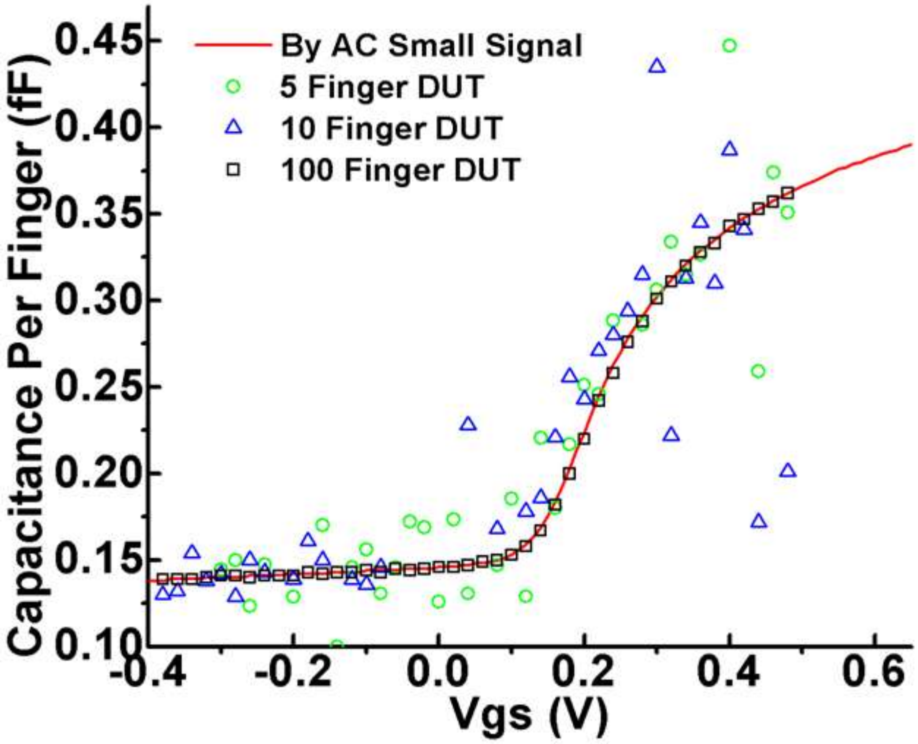


**Fig. 3-8**: Bias-dependent capacitance derived by the CBCM setup in Fig. 3-4 (c) with 3, 5, and 100 parallel DUT fingers. Solid line is the exact capacitance calculated by AC small signal method in MEDICI. Large errors are observed for DUTs with 5 or lesser no. of fingers.

The setup (b) in **Fig. 3-4** with pass-gates to reduce impact of charge injection offers the best accuracy even with a 3 finger DUT (which represents a total DUT capacitance of ~1.35fF). The setup (c) in **Fig. 3-4** is accurate for large capacitance with the accuracy degrading for smaller capacitances.

### *3.4 Analysis of Main Sources of Errors*

During the switching on and off process of the pseudo-inverters, the charge imbalance at the gates of N and P Drivers results in feed-forward current being injected to $V_{DD}$ or $V_{SS}$, contributing a third component $C_{inj}$ in the measured total capacitance: $C_{meas}=C_{par}+C_{DUT}+C_{inj.}$ In order to extract $C_{DUT}$, a baseline value $C'_{meas}=C'_{par}+C'_{inj}$ is measured by means of a matched unloaded pseudo-inverter, which leads to:

$$C_{DUT}= C_{meas} -C'_{meas}= C_{par}+C_{DUT}+C_{inj}- C'_{par}-C'_{inj} \qquad (3\text{-}5)$$

The accuracy of the measurement is limited by the mismatch between the baseline value (provided by an unloaded pseudo-inverter) and the parasitic and change injection contribution measured by the DUT test circuit (provide by a loaded pseudo-inverter).

For charge injection during measurement, the amount of charge injected to $V_{DD}$ depends on the rise/fall times of the pulses, magnitude of the $C_{gs}/C_{gd}$ and also on the load at node *X*. Because the load is inherently different while measurement of $C_{meas}$ and $C'_{meas}$ are being made, there will be a systematic error $C_{inj}-C'_{inj}$.

On the other hand, $C_{par}$ may not equal $C'_{par}$ due to layout mismatches (which can be avoided by tactful lithography design) and process variations. Since process variation is more difficult to control, we can consider $C_{par}-C'_{par}$ a source of random noise. The instrumental noise and limited resolution of current measurement also contribute to this random noise.

A third source of error is the numerical error introduced by the differentiation process as in **Eqn. (3-4)**.

We will look into each of these sources of error in this section.

### *3.4.1 Charge Injection Error*

The effect of systematic charge injection error $C_{inj}$-$C'_{inj}$ is illustrated further in **Fig. 3-9** using setup (c) of **Fig. 3-4**. $C'_{meas}$, obtained by differentiating charging current $I'_{V_{DD}}$ for 100 finger and a single finger DUT, are compared with the parasitic capacitance given by the compact model (BSIM3) of the driver devices.

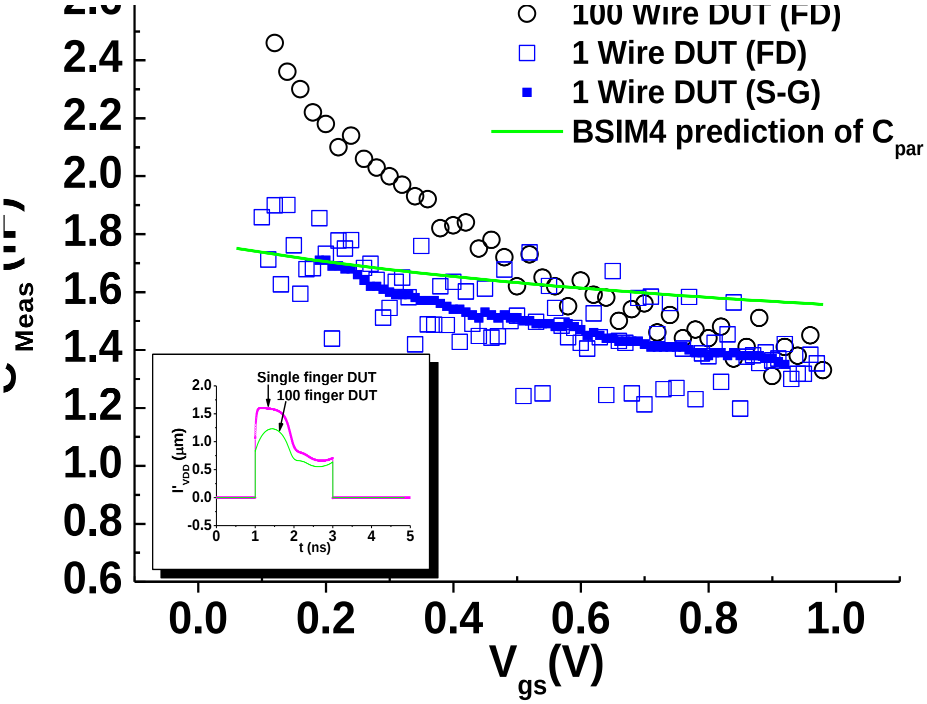


**Fig. 3-9**: Symbols are $C'_{meas}$ derived by CBCM simulation for 100- and 1-finger DUT using FD, and for 1-finger DUT derived using S-G method. Solid line shows the parasitic at node X obtained from compact model (BSIM3) for the drivers used in the mixed-mode simulation. The discrepancy is due to charge injection and numerical errors. Inset: The difference in time dependent current $I'_{V_{DD}}$ for DUT with 100-finger and a single-finger due to effects of charge injection. Although the DUT is invisible during the charging cycle in both cases, magnitude of its capacitance still affects the amount of charge injected to $V_{DD}$.

It is worth noting here that although the DUT is supposed to be “invisible” during measurement of current $I'_{V_{DD}}$ as its source/drain terminals have been pre-charged to the same voltage as that at node X, its presence still affects the charge injection. This is because during the transition period in which the drivers are being turned on or off, the net voltage between DUT gate (node X) and S/D terminals is non-zero and varies with time. The displacement current (C $d$V/$d$t) influences the charge flowing from/to $V_{DD}$ ($V_{SS}$). Therefore, $C_{DUT}$ still plays a role in deciding the amount of charge injected to $V_{DD}$. This is illustrated in the inset to **Fig. 3-9** which shows a clear difference in magnitude of time-dependent currents $I'_{V_{DD}}$ for 100-finger and 1-finger devices and explains the difference in the parasitic calculations from BSIM3 models and CBCM computations as well as the difference between parasitic computed for 100-finger and 1-finger DUTs with the same set of drivers. Therefore, to ensure the systematic error due to charge injection is less, $C_{DUT}$ can not be very large in magnitude comparing with $C_{meas}$, so that the mismatch in loaded and unloaded test circuit is too large.

### ***3.4.2 <u>Random noise due to mismatch/variation</u>***

The second source of error is the random noise coming from mismatch of device dimension and properties as the width of the gate-to-source/drain overlap region in the DUT branch and reference branch due to process variation. Setup (c) in **Fig. 3-4** avoids this problem by performing the two measurements on the same circuit. On the other hand, although current can be tapped and integrated with much accuracy in simulation, in actual measurements, the instrument that measures the average current will inevitably introduce inaccuracy, adding to the random noise in measurement. Therefore, we go a step further to evaluate these three setups when random error of a specific magnitude range is introduced to the average current.  The measurement accuracy is typically specified as a fraction of the measured value, e.g., 0.5% of the measured current value. To emulate the effect of measurement inaccuracy, we introduced a random noise up to a fraction of 1% in the integrated charge. In **Fig. 3-10**, the RMS error as a percentage of the total capacitance extracted is plotted against the maximum of the introduced random noise in percentage. It shows that in order to measure the low values of capacitance accurately, the average currents in the CBCM methods should be measured to 0.1% accuracy or better. This requirement would progressively relax for relatively higher values of DUT capacitance.

**Fig. 3-10** also shows that the three setups have different degree of tolerance towards the random error. The setup (c) appears to be most susceptible to random noise that a 0.2% measurement error would result in more than 30% RMS error for low capacitance DUT. In the later section we will show that this error, together with the random noise due to numerical error can be corrected by Savitzky and Golay's (S-G) algorithm [16].

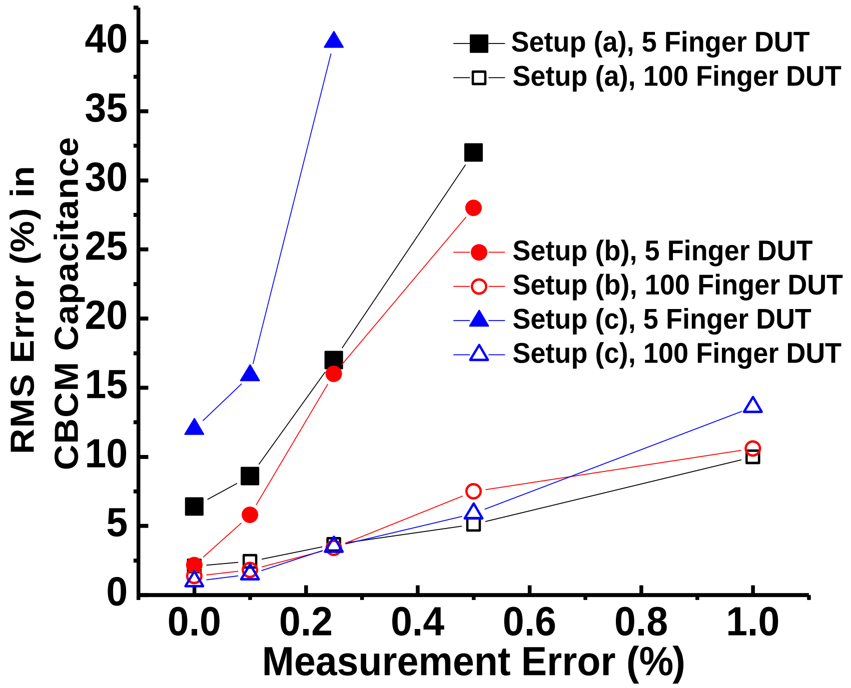


**Fig. 3-10**: RMS error plotted against introduced random error in the charge for setups of Fig. 3-3 (a)-(c). Large DUT capacitance enhances tolerance to error in all the methods.

### *3.4.3* ***Random noise due to numerical error***

The third source of error is introduced by the numerical evaluation of the capacitance using **Eqn. (3-4)**. The numerical inaccuracy becomes especially prominent when the DUT capacitances become comparable to or lower than the parasitic capacitance presented by the pseudo-inverters. For very small DUT, charging currents $I_{V_{DD}}$ and $I'_{V_{DD}}$ become very close in magnitude; therefore any error in the magnitude of the measured current gets amplified during subtraction and differentiation process, resulting in noisy $C_{DUT}$-$\mathrm{V}_{\mathrm{gs}}$ curves.

Both being random noises, the mismatch error and numerical error in differentiation can be corrected by following S-G algorism. The S-G method performs a local polynomial regression on a series of equally spaced points to determine the smoothed value for each point. The advantage of this method is that it is capable of preserving the features of the distribution such as relative maxima, minima and width, which are usually flattened by other adjacent averaging techniques (e.g., moving average). The same method can also be applied to calculate the derivatives of the smoothened data. As shown in **Fig. 3-9**, the noise in $C_{par}$ is reduced substantially using S-G differentiation algorithm as compared with conventional finite difference (FD) method using central difference formulae.

The simulation results in terms of per finger capacitance for DUTs having different number of nano-wire fingers are shown in **Fig. 3-11**, and are compared with the capacitance obtained from small signal analysis in MEDICI using signal level of (kT/10q) at 1MHz. It can be seen that the $C_{DUT}$ simulated by CBCM is smoother with higher number of fingers in the DUT, i.e., higher total DUT capacitance value. Comparing the two curves for 5-finger DUT using FD method and S-G method, it is clear that S-G algorithm effectively reduces the fluctuations and obtains smoother C-V curves. If S-G algorithm is properly applied, even single finger DUT device whose magnitude is less than 380 aF can be extracted/measured with reasonable accuracy.

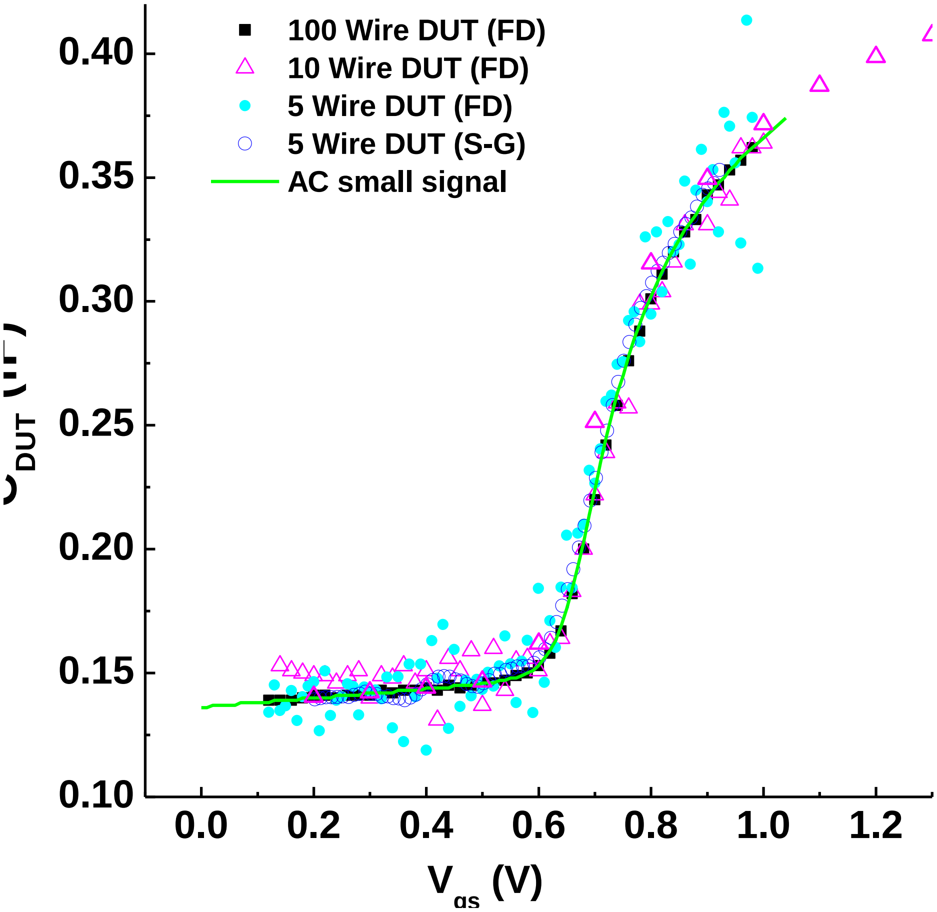


**Fig. 3-11**: C-V curves obtained by CBCM method for DUT with 100, 10 and 5 fingers using FD and 5-finger $C_{DUT}$ curve using S-G method. Solid line shows the C-V characteristics obtained for the nano-wire device using small-signal analysis (Signal level= (kT/10q) and frequency: 1MHz).

Moreover, it was found that S-G method is effective in correcting the charge injection induced error as well. As shown in **Fig. 3-9**, the $C'_{meas}$ data obtained using S-G algorithm to do differentiation is much smoother than that derived by conventional finite difference method.

## *3.5 Assessment of CBCM Efficacy and Its Limits*

In order to properly design the CBCM test-key, it is important to understand the lower limit of the CBCM method and the factors affecting it. Although, as mentioned above, the frequency of the pulses should not affect the extracted result of $C_{DUT}$ if both the parasitic and DUT capacitances are very small, $I_{V_{DD}}$ and $I'_{V_{DD}}$ might both be too low to be assessed accurately. This will cause error in the extracted DUT capacitance on the account of degraded accuracy of measured current. In such cases, increasing the pulse repetition frequency will increase the current through $V_{DD}$ and in turn improve the accuracy of the technique. In terms of test-key design, the most important factor is the ratio of $C_{DUT}$ with respect to $C_{par}$.

From the discussion of the previous section, it is not difficult to see that to reduce the errors of capacitance extraction in CBCM, a compromise is needed among sources of errors:

For larger CDUT relative to $C_{par}$, or $C_{meas}$, the systematic error due to charge injection ($C_{inj}$-$C_{inj0}$) increase due to a larger difference in load in the reference circuit and DUT test circuit. For smaller $C_{DUT}$ relative to $C_{par}$, or $C_{meas}$, the random noise due to process variation and numerical error would be more significant, and also leads to loss of accuracy.

To illustrate this point further, we plot the RMS error in CBCM capacitance against the ratio $C_{DUT}/C_{par}$ in **Fig. 3-12** ased on simulation result using setup (c) in **Fig. 3-4** For $C_{DUT}$ values less than half the $C_{par}$ values, the error is less than 5% using FD method and improves to 3% using S-G algorithm. The RMS error increases rapidly as the value of $C_{DUT}$ falls below 20% of $C_{par}$ as can be seen in the inset of **Fig. 3-12**- the error increases by more than 10% for each 5% increase in $C_{par}$. The simulations bring out the fact that the parasitic capacitance due to the drivers of pseudo-inverters and interconnects should be less than twice the minimum capacitance to be measured using CBCM to ensure the measurement accuracy of better than 5%.

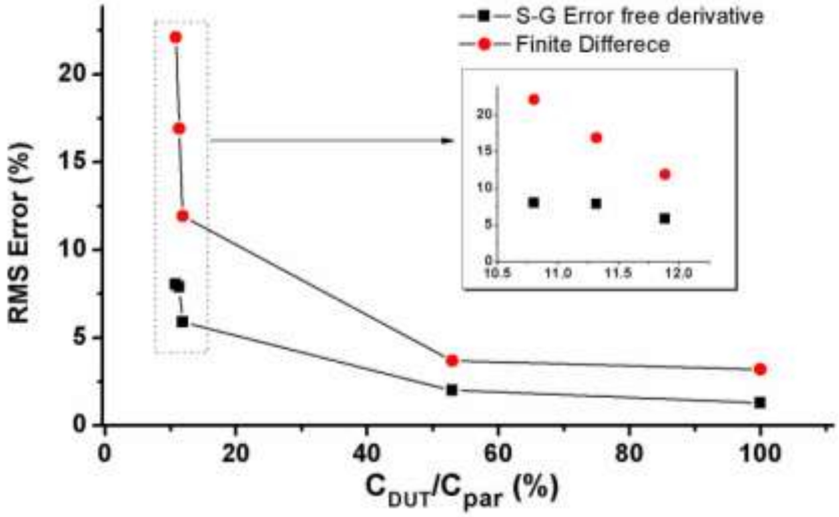


**Fig. 3-12** RMS error over the bias range considered in the CBCM simulations plotted against the relative ratio of $C_{DUT}$ and $C_{par}$. The error increases rapidly when $C_{DUT}$ is less than 20% of the magnitude of $C_{par}$. Inset shows the details for $C_{DUT}/C_{par}$<12%.

### *3.6 Chapter summary and conclusion*

In this chapter, we introduced and evaluated different CBCM setups for low capacitance measurement in detail for sub-femto farad voltage-dependant capacitance measurement using intensive mixed-mode TCAD simulations. The effects of charge injection and numerical error are delineated. Practical design guidelines are arrived at for the desirable relative size of $C_{par}$ and $C_{DUT}$. Noise free differentiation scheme proposed by Savitzky and Golay was found to give more accurate and smooth derivatives for CBCM capacitance extraction.

Overall, we find setup (c) in **Fig. 3-4** to be more suitable for fF scale capacitance because:

i. Mismatch and process variation is a significant and inevitable problem with nanoscale device fabrication. Setup (c) addresses this problem successfully by performing the DUT and reference measurement on the very same physical test structure.

ii. Charge injection induced systematic error has been reduced by having the DUT physically connected while performing the reference measurement.

iii. Efficacy can be improved by optimizing ratio of $C_{par}$ and $C_{DUT}$.

iv. Simple layout and realization.

These detail simulation study not only provide understanding of CBCM mechanism and accuracy range, it also gives design guidelines for practical application of CBCM. CBCM has been shown to be a promising characterization technique for fF or even lower range capacitance measurement without the complication of advance equipment, shielding and high noise level in many other approaches. These results motivate us to explore it further in experiments so that the result can be applied in modeling and calculation of carrier mobilities. In the following three chapters, we will report the fabrication, characterization and analysis nanowire capacitance using CBCM.

## *Reference for Chapter 3*

[1] "Emerging Research Device", ITRS 2006

[2] H. Sakaki, "Scattering suppressed and high-mobility effect of size-quantized electrons in ultrafine semiconductor wire structures", *Jpn. J. Appl. Phys*., vol.19, no. 12, pp. L735-738, 1980.

[3] Y.Cui, Z. Zhong, D. Wang, W. U. Wang, and C. M. Lieber, "High performance silicon nanowire field effect transistor", *Nano Lett*., vol. 3, no. 2, pp.149-152, 2003.

[4] R. Kotlyer, B. Obradovic, P. Matagne, M. Stettle, and M. D. Giles, "Assessment of room-temperature phonon-limited mobility in gated silicon nanowires", *Appl. Phys. Lett.*, vol. 84, no. 25, pp. 5270-5272, 2004.

[5] Edwin B. Ramayya, D. Vasileska, S. M. Goodnick, and I. Knezevic, "Electron Mobility in Silicon Nanowires", *IEEE Trans. Nanotech.*, vol. 6, no. 1, pp. 113-117, 2007

[6] J. Hattori, S. Uno, N. Mori, and K. Nakazato, "A theoretical study of electron mobility reduction due to acoustic phonon modulation in a free-standing semiconductor nanowire", *Tech. Dig. ESSDERC 2008*, pp. 289-292.

[7] S. Poli, M. G. Pala, T. Poiroux, S. Deleonibus, and G. Baccarani, "Size dependence of surface-roughness-limited mobility in silicon-nanowire FETs", *IEEE Trans. Elec. Dev.*, vol. 55, no. 11, pp.2968-2976, 2008.

[8] Jiezhi Chen, T. Saraya., Kousuke Miyaji, K. Shimizu and T. Hiramoto, "Experimental study of mobility in [110]- and [100]-directed multiple silicon

nanowire GAA MOSFETs on (100) SOI", Tech. Dig. VLSI Technology Symposium, 2008, pp. 32-33.

[9] T. Tezuka, E. Toyada, S. Nakaharai, T. Irisawa, N. Hirashita, Y. Moriyama, N. Sugiyama, N. Taoka, Y. Yamashita, O. Kiso, M. Harada, T. Yamamoto, and S. Takagi, "Observation of mobility enhancement in strained Si and SiGe Tri-Gate MOSFETs with Multi-Nanowire channels trimmed by hydrogen thermal etching", *Tech. Dig. IEDM 2007,* pp. 371-374.

[10] K. E. Moselund, P. Dobrosz, S. Olsen, V. Pott, L. De Michielis, D. Tsamados, D. Douvet, A. O'Neill, and A. M. Ionescu, "Bended Gate-All-Around nanowire MOSFET: a device with enhanced carrier mobility due to oxidation-induced tensile stress", *Tech. Dig. IEDM 2007,* pp. 191-194.

[11] B. Laquai, H. Richter, and B. Hofflinger, "A new method and test structure for easy determination of femto-farad on-chip capacitances in a MOS process", *Proc. IEEE Int. Conf. on Microelectronic Test Structures, 1992,* pp. 62-66.

[12] Sell et al., "Charge-based capacitance measurements (CBCM) on MOS devices", *IEEE Trans. Device and Material Reliability*, vol. 2, no. 1, pp.9-12, 2002.

[13] J.C. Chen, B.W. McGaughy, D. Sylvester, and C. Hu, "An on-chip, attofarad interconnect charge-based capacitance measurement (CBCM) technique", Proceedings of IEDM '96, pp.69-72B.

[14] L. Vendrame *et al.,* Proc. of the 7th IEEE SPI Workshop, 2002.

[15] Y. W. Chang, H.-W. Chang, C.-H. Hsieh, H.-C. Lai, T.-C. Lu, W. Ting, J. Ku, and C.-Y. Lu, "A novel simple CBCM method free from charge injection-induced errors", IEEE Elec. Dev. Lett. 25(5), pp.262-264, 2004.

[16] A. Savitzky and M. J. E. Golay, "Smoothing and Differentiation of Data by Simplified Least Squares Procedures". Analytical Chemistry vol. 36 no. 8, pp. 1627–1639, 1964.

# Chapter 4

# Fabrication of Si Nanowire Devices

## *4.1 Introduction and Si NW Process Flow Overview*

The fabrication technology of Silicon Nanowire (SNW) device channels can be broadly categorized into two groups, namely, the bottom-up approach and the top-down approach. In case of bottom-up process, the nanowire channels are synthesized using catalyst growth or chemical-vapor-deposition [1] and using stress-assembly technologies [2]-[4]. Other forms of bottom-up NW fabrication techniques includes etching with platinum masks prepared by superlattice NW pattern transfer (SNAP) process [5][6] and templated growth [7]. However, to make the NW channels into functional electron devices, effective and efficient integration and assembly techniques must be developed to transfer the nanowires from growth substrate onto the device substrate. A lot of technical challenges still exist in the integration of nanowire [8][9]: controlled orientation and spatial positioning in wafer scale is extremely difficult; hierarchical organization of NW channels on multiple length scale adds to the complexity of integration; and making interconnects between the nano-structured device and micro-structured parts of the circuitry such as metal line connections and bonding pads remain challenging as well.

The top-down fabrication approach, on the other hand, takes advantage of the existing CMOS technology platform. All variants of top-down process start with standard Silicon or Silicon-on-Insulator (SOI) wafer and involve lithography and/or etching process to form a starting fin pattern. Then the fin is reduced to device channel with nanometer diameter by hardmask etching/trimming [10], selective SiGe removal [11]or stress-limited oxidation [12]. A common and most significant advantage of top-down approaches is the easy integration in conventional CMOS platform, making it readily suitable for large scale integrated circuit implementation. The main challenge faced by this technique is that the diameter of the SNW channel is not easily controlled, and the process variation is manifested quite significantly in fabricated nanoscale devices.

The Silicon Process Technology department of Institute of Microelectronics has extensively developed and implemented stress-limiting oxidation process for SNW fabrication on 8-inch CMOS process line for a few years. Therefore, this process is the natural choice for our experiment, leveraging on the existing fabrication expertise. **Fig. 4-1** illustrates the process flow of fabrication by schematics.

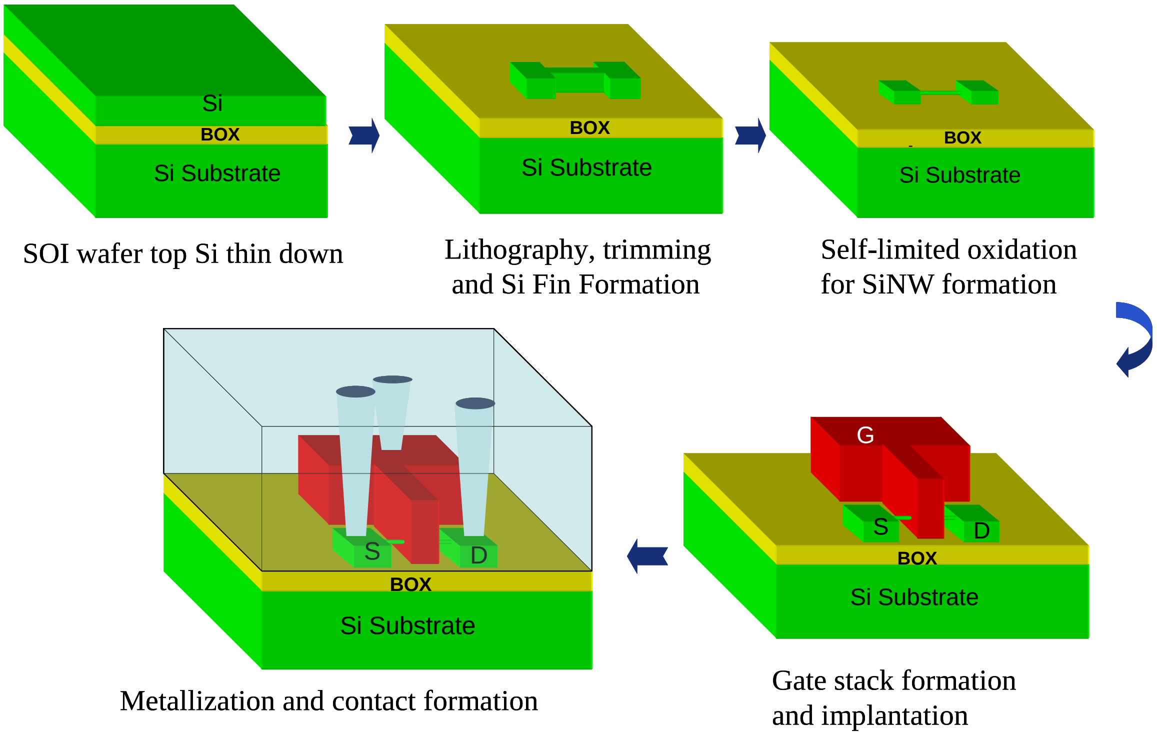

Si
BOX
Si Substrate
SOI wafer top Si thin down
BOX
Si Substrate
Lithography, trimming and Si Fin Formation
BOX
Si Substrate
Self-limited oxidation for SiNW formation
G
S
D
BOX
Si Substrate
Gate stack formation and implantation
S
D
BOX
Si Substrate
Metallization and contact formation

**Fig. 4-1**: Process flow and schematic diagrams for SNW device fabrication

First, the SOI wafer is thinned down by dry oxidation for optimized fin shape. Then the fin and source/drain contact region pattern were transferred to silicon by lithography and etching. The fin was oxidized in dry oxide so that its diameter reduces to the nano meter level, forming the channel of the nanowire device. After that, gate stack is formed and dopants were implanted as in the conventional CMOS process. This is followed by re-metal dielectric deposition and contacts etch. Finally, metal was deposited, etched and alloyed, completing the fabrication process. Although this process flow is adapted from a well established CMOS technology, there are still issues specific to the nanoscale three dimensional devices such as the poly stringer problem which increases the parasitic capacitance of the devices quite significantly. We have modified the processes and added in new steps in order to remove the topology and the stringer. Also, special masks have been designed for CBCM testkey for measurement of sub-fF scale capacitance. The fabrication technology for stringer-free SNW device will be discussed in detail in this chapter.

## 4.2 Substrate Preparation, Lithography and Fin Formation

In order to realize both N and P-type NW devices on the same SOI wafer, the process involves at least 6 mask layers, much like the conventional CMOS process. These layers are: active [Fin (Nanowire) , Source and Drain definition] layer, Poly-Si gate layer, contact layer, metal layer, as well as N and P-type dopant implant layers, as shown in the schematic layout shown in **Fig. 4-2**.

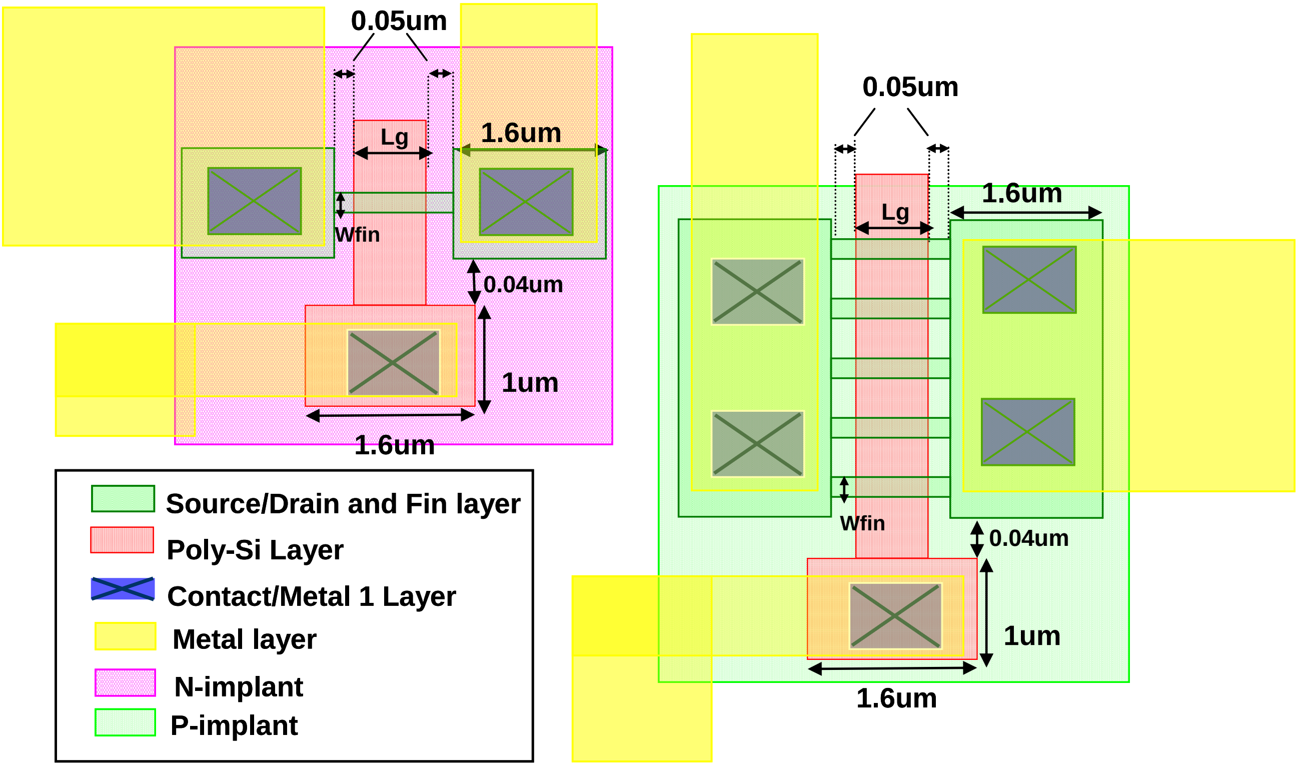


**Fig. 4-2**: Mask layout design schematic for both N and P type single and multiple finger NW devices.

The channel region is formed by a fin structure connected to two larger pads for Source and Drain (S&D) contacts formed by etching. A hanging nanowire hinged on S&D pads will then be formed by stress-limited oxidation which requires well-controlled dimensions of the fin. Our simulation and short loop experimental results have shown that the most optimized fin cross section dimension for effective stress-limited oxidation is 50 nm × 50 nm. The thickness of the fin is defined by the thickness of the silicon layer on which the fin is patterned. Thermal oxide growth is used to thin down the initial silicon layer from 70 nm to 50 nm. On the other hand, the width of the fin is defined by the lithography. Using the 248 nm optical lithography system available to us, the minimum feature dimension that can be achieved with good line integrity is around 150 nm. To reduce the fin width further, the resist is trimmed to 50 nm using $O_2$ based plasma etchant. Then the top silicon layer is etched to form individual Si islands on which device will be formed.

One of the major problems faced by the lithography scheme described above is the “optical proximity effect” [13] due to the very small feature size compared with the wavelength of the light in the lithography system. This effect manifests itself most significantly in the area in which the fin is joining the S&D contact region – although they are designed to be perpendicular to each other with a 90° corner, the developed pattern shows a flaring across it approximately 75 nm wide as shown in **Fig. 4-3** (a). Even though the flaring region does not affect the device in operation, it is highly undesirable due to the following reasons: first, its presence limits the capability of gate length scaling, control of device dimension becomes more difficult for shorter gate lengths. Second, due to the wider cross section, this flaring region will not convert to nanowire as other parts of the fin. Thus it will form a partially gate-all-around channel with the two ends having tri-gate structure.  Thirdly, its non-uniform shape and dimensions of the channel region adds difficulty to device characterization and modeling.

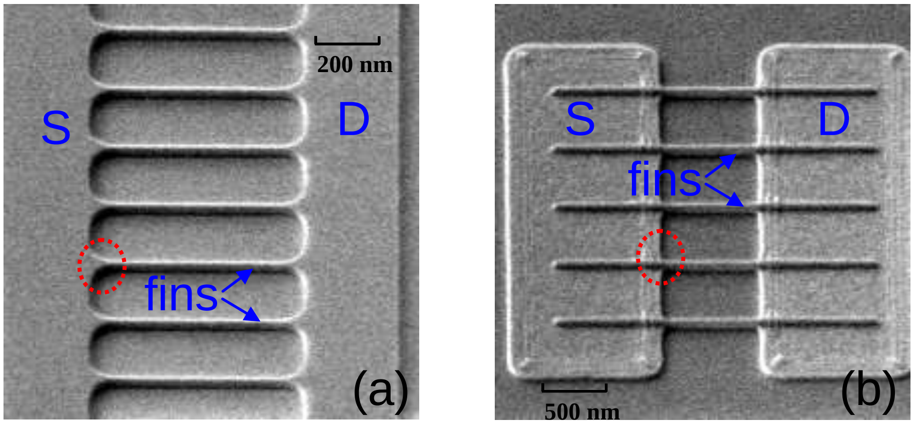


**Fig. 4-3**: Images of the NW device structure after fin and S&D etch (a): using single mask lithography, circled region showing the flaring at the corner of each ends of the fin (b): using double mask two step lithography, circle region showing that the flaring has been eliminated

To solve the flaring problem, we adopt a two masks, two steps approach for the fin and S&D definition. The process is illustrated by schematics in **Fig. 4-4**. The fin pattern without S&D is first transferred to a hard mask of 100 Å Silicon Nitride during the 1st lithography process. During the 2nd lithography, the S&D pattern is transferred to photo resist. Then both patterns are transferred together to Si layer by dry etch using photo resist and SiN as mask. As shown in Fig. 4-3 (b), the proximity effect is eliminated using this method. The hard mask can be removed by wet etch after pattern transfer.

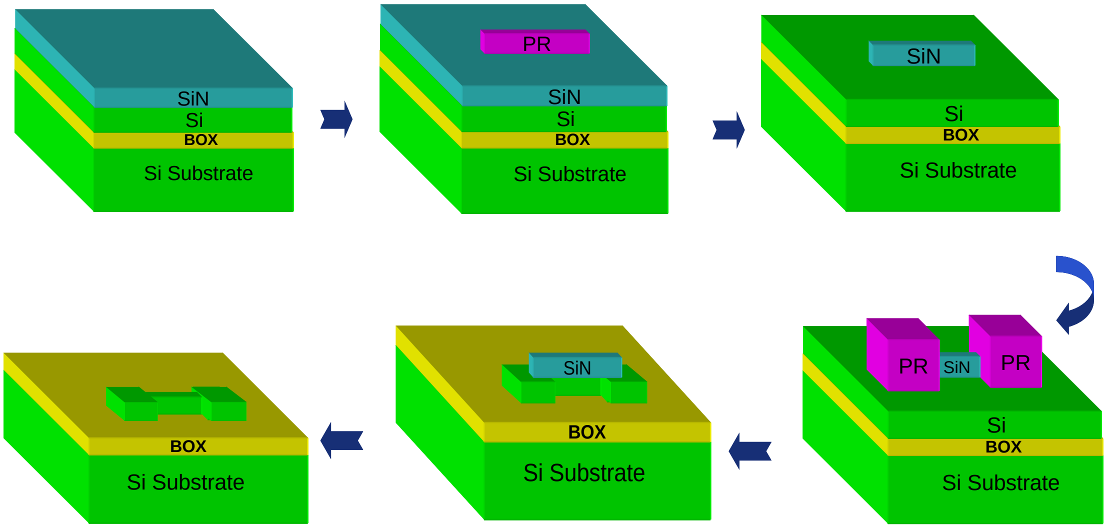


**Fig. 4-4**: Schematic illustrating the two masks, two steps process for fin and S&D definition to eliminate the proximity effect.

## 4.3 Stress-limited Oxidation for Nanowire Formation

To study the characteristics of silicon nanowires, Liu et al. reported a novel technique to produce vertical silicon nanowires with a diameter down to 2nm in 1993 [14]. By subjecting the vertical Si pillars to dry oxidation, they found that the oxidation rate reduced significantly after a thick oxide is formed. Depending on oxidation duration, silicon nanowires with a large range of diameters were then available. They claimed to observe stress-limiting phenomenon with dry oxidation temperature no more than 875ºC. With stress-limiting effect, the diameter of silicon core saturates in the range of 2nm to 10nm depending on the starting nanowire diameters and dry oxidation temperature. The relationship due to stress-limiting effect under different temperature conditions is summarized in Fig. 4-5.

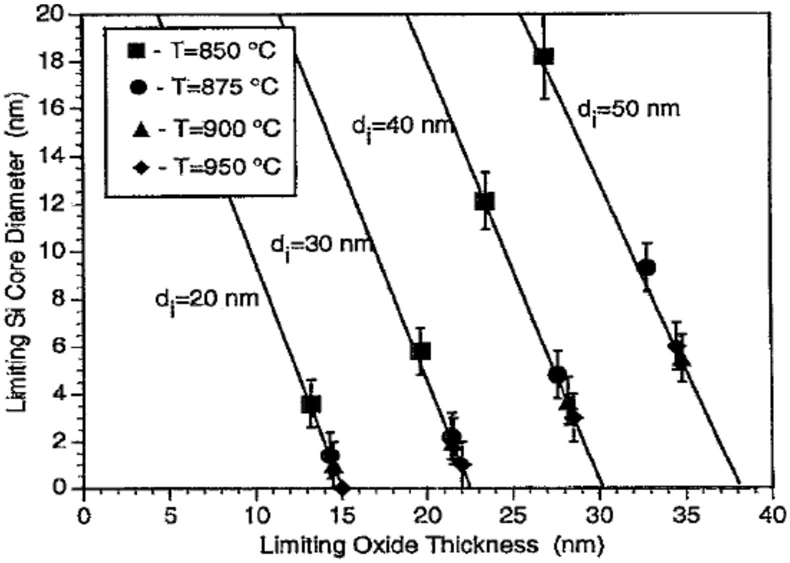


**Fig. 4-5**: The relationship between limiting Si core diameters and the corresponding limiting column oxide thicknesses when the core reduction rate is reduced (less than 0.2 nm/h). Each group of data points on a single extrapolated straight line corresponds to an identical starting column diameter, di. As the oxidation temperature increases from 850 to 950 ºC, the data points move toward a smaller core diameter and a larger oxide thickness [14].

Liu et al. [12] argues that the origin of the stress-limiting phenomena is kinetically limited oxidation reaction at Si column surface and the asymptotic behavior arises from accumulated stress. Therefore, the temperature and time at which stress-limiting happens depends on a number of parameters including the interface reaction rate which is also a function of temperature and stress, the interface stress which is related to both Si core curvature and the oxide thickness, and $O_2$ diffusivity in oxide which decreases by compressive stress and increases by tensile stress, etc.

Due to the complexity of the problem and lack of well-calibrated stress model, the TCAD simulation tool TSUPREM4 available to us does not predict stress-limiting phenomena using default model. Simulation and several short loop experiments were carried out to study formation of NW channel by oxidation.

### ***4.3.1 2D Diffusion/oxidation model***

For 2D silicon oxidation, there are four numerical models: Vertical, Compress, Viscous and ViscoEla model. Due to 2D nature of this simulation, Viscous or ViscoELA must be selected to ensure proper stress calculation. The Viscous model simulates viscous flow of the oxide during oxidation using 7-node finite elements, which allow accurate values of stress to be computed. The ViscoELA model can also calculate the viscoelastic flow of the oxide by using 3-node finite elements, which use numerical techniques that allow approximate values of stress to be computed. Correct stress estimation in our simulation is very important. However, the Viscous model suffers great convergence problems and low speed for structures with a larger number of nodes. The ViscoELA model was selected throughout all simulation runs.

In 2D numerical oxidation, oxidant diffusivity is calculated in oxide. Estimated surface reaction rate is applied at the Si/$SiO_2$ interface.

The diffusivity is calculated from the parabolic oxidation rate B in Eqn. 4-1.

$$D = B\frac{N_1}{2C^*} \quad = B\frac{N_1}{2HP_0} \quad = B\frac{N_1}{2H} \tag{4-1}$$

$C^*$ is equal to Henry's law coefficient given by HENRY.CO at 1atm. $N_1$ is given by,

$$N_1 = \begin{cases} THETA & \text{for } O_2 \\ 2 \times THETA & \text{for } H_2O \end{cases} \tag{4-2}$$

The surface reaction rate is derived from,

$$k_s = \frac{B}{A}\frac{N_1}{C^*} \tag{4-3}$$

The zero stress viscosity is material dependent,

$$\mu = Visc.0\exp\left(\frac{-Visc.E}{kT}\right) \tag{4-4}$$

When stress is taken into consideration in 2D oxidation, the surface reaction rate $k_s$, diffusivity of oxidant in oxide $D$, and the oxide viscosity $\mu$ are modified to reflect their dependence on the stresses in the oxide:

$$k_s^{'} = k_s \exp\left(-\frac{\sigma_n VR}{kT}\right) \tag{4-5}$$

$$\mathrm{D}^{'} = D\exp\left(-\frac{pVD}{kT}\right) \tag{4-6}$$

$$\mu = \mu_0 \frac{\sigma_s VC/2kT}{\sinh(\sigma_s VC/2kT)} \tag{4-7}$$

$k_s^{'}$, $D^{'}$ and $\mu$ are the stress-dependent reaction rate, diffusivity, and viscosity, respectively. *VR*, *VD* and *VC* are parameters on the AMBIENT statement. The exponential term in Eqn. 4-6 is limited for positive arguments (i.e., the value of $p$ must be negative) to the value of VDLIM (a parameter on the AMBIENT statement) to prevent unrealistic enhancement of the diffusivity. The surface reaction rate depends both on the stress normal to the interface:

$$\sigma_n = -(\sigma_{xx}n_x^2 + \sigma_{yy}n_y^2 + \sigma_{xy}n_x n_y) \tag{4-8}$$

Where $n_x$ and $n_y$ are the components of the unit vector along and normal to the interface. The oxidant diffusivity depends on the hydrostatic pressure $p$ defined by:

$$p = -\frac{1}{2}\left(\sigma_{xx} + \sigma_{yy}\right) \tag{4-9}$$

The oxide viscosity depends on the total shear stress $\sigma_S$:

$$\sigma_s = \sqrt{\frac{1}{4}\left(\sigma_{xx} - \sigma_{yy}\right)^2 + \sigma_{xy}^2} \tag{4-10}$$

### ***4.3.2 Shape control of the NW channel***

The shape of the NW channel formed by oxidation is found closely related to the shape and size of the initial fin as well as the oxidation temperature. A tall rectangular shaped fin results in two triangle shape of nanowire as demonstrated by both simulation and experiments. As illustrated in **Fig. 4-6**, the initial fin was matched with the experimental device which has a width of 56nm and the initial boron doping is $10^{15}$ cm$^{-3}$. Its top surface is (100) while its side surface is (110). Oxidation was carried out in dry $O_2$ ambient of 875ºC.

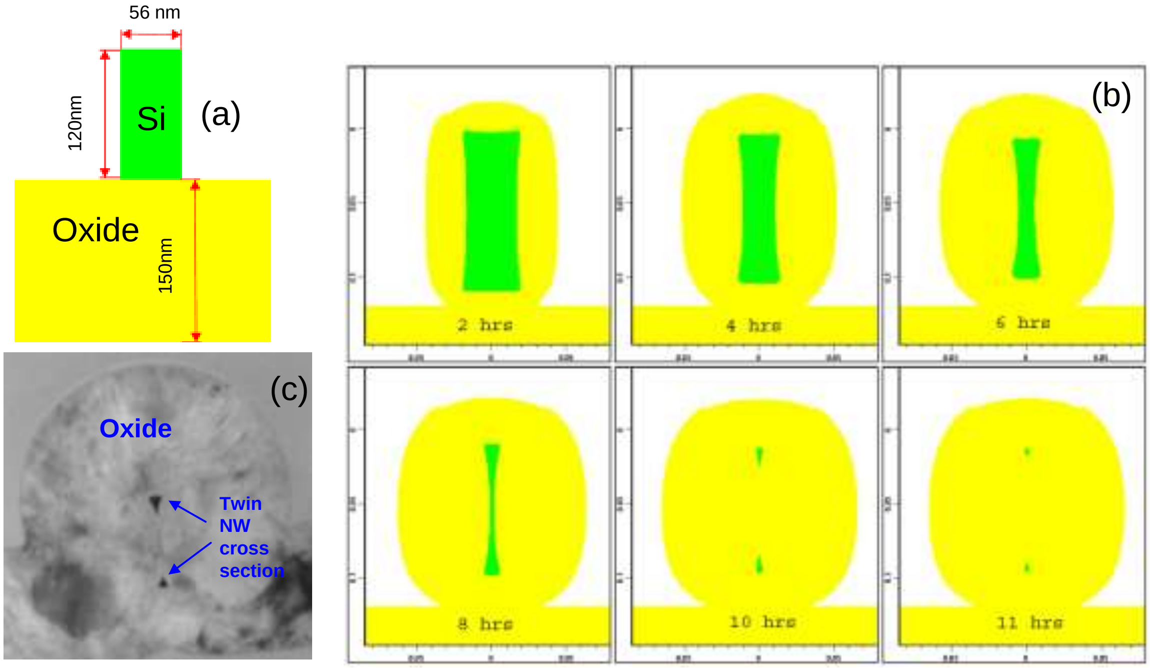


**Fig. 4-6**: (a)- the simulation structure cross section (b)-the silicon fin profile evolution under dry oxidation at 875ºC. (c)-the experiment TEM cross section of twin triangular NW

As indicated in **Eqn. 4-9**, compressive stress in $SiO_2$ hinders diffusion of oxidants while tensile stress promotes it. There are two obvious findings from the hydrostatic pressure contour map extracted from simulation. First, the tensile stress is observed at $Si/SiO_2$ interface around the center of the silicon fin after 60 minutes into the oxidation process. Then the stress near the Si surface gradually changes into compressive. The initial ample oxidants supply around the center of the silicon fin helps increase oxidation rate locally, which leads the center pinch-off eventually. Secondly, the four corners of the silicon fin suffer high compressive stress all along oxidation process. Their oxidation rate is much slower than the other part in the structure. Therefore the twin triangular NW shape forms. However, it is remarkable that though the stress effect on shape is significant, it does not results in a stress-limiting behavior. In our experiment, the Si wire is fully consumed after 4.5 hrs of oxidation.

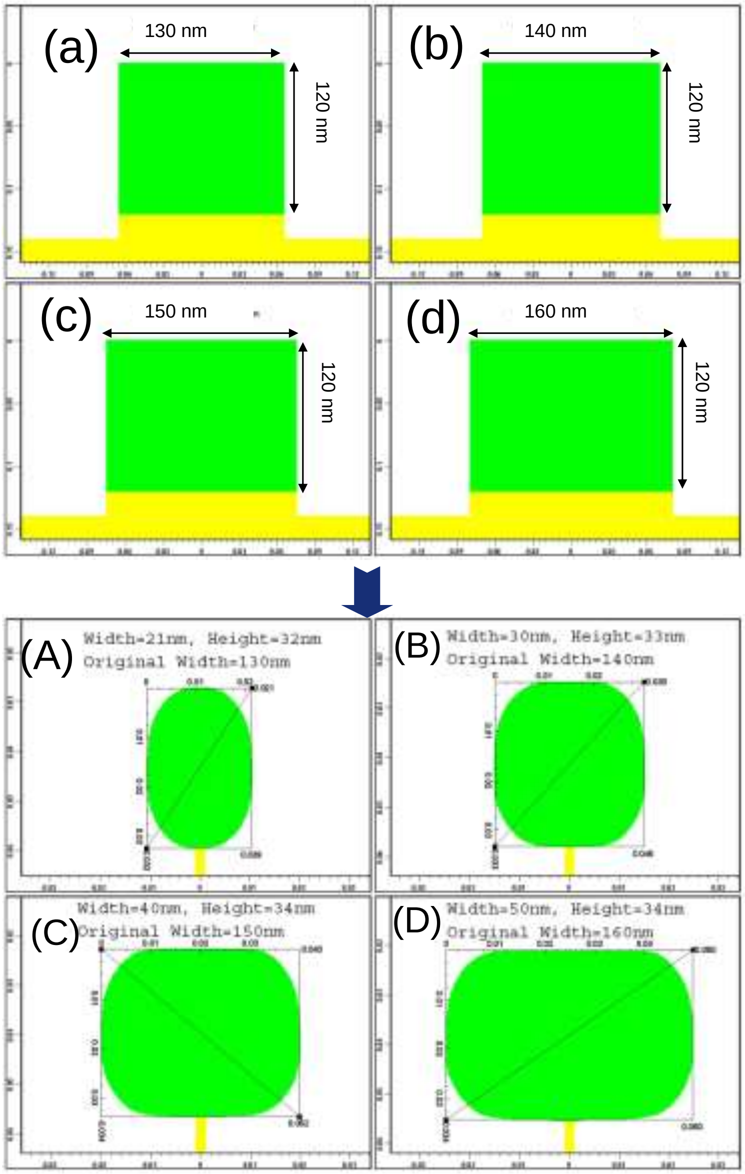


**Fig 4-7**. (a)-(d):The initial structure has a variation of 130~160nm in width. (A)-(D): the corresponding profiles of the silicon cores after 1150ºC 20 minutes oxidation.

It is also found that the nearly square fin shapes tend to be better rounded under higher oxidation temperature due to oxide reflow and release of stress as illustrated in **Fig. 4-7** (a)-(d)**.** It shows that oxide on the side walls of the fin oxidizes faster under high temperature condition. Therefore in order to reduce the fin cross section from rectangular to circular, the best starting shape is a squarish fin whose width is slightly wider than the height (~20 nm in the case of Fig. 4-7(c)). However, the drawbacks of using high temperature oxidation are many: (1) the oxidation rate is very high, making process control difficult and uniformity poor (2) the reflow releases stress in oxide so that the process cannot be stress-limited. Therefore, we adopted a process which combines a short high temperature cycle and a longer low temperature cycle. In this way, the shape rounding can be achieved and it provides much better control of process. A TEM image of the sample cross section is shown in **Fig. 4-8**.

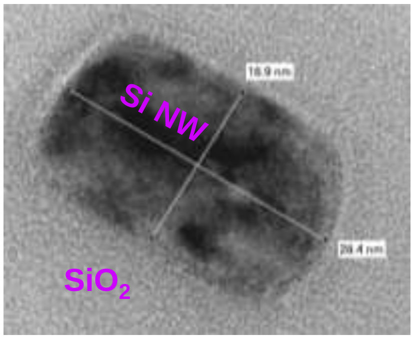


**Fig. 4-8**: TEM picture of the silicon core profile after 1050ºC 5 minutes and 875ºC 4 hours dry oxidation. Good corner rounding is observed. The silicon core profile is elliptic. Estimate dimension of the starting fin is: width=70 nm height=50nm

### *4.3.3 Observation of stress-limited oxidation*

Although the 2D simulation program SUPREM4 does not predict stress-limited oxidation, clear evidence of stress-limiting effects has been found in our experiments. A batch of suspending fin structure with rectangle cross section (height≈50nm, width≈35nm) was fabricated and subjected to different oxidation conditions. Extensive TEM study has been carried out for the oxide thickness and the Si core dimension. Shown in **Fig. 4-9** (a) is the cross section TEM image of the initial fin before oxidation. **Fig. 4-9** (b) shows the cross section image after the sample was subject to 975ºC dry $O_2$ oxidation for 80 min. The Si core dimension has been reduced by two third suggesting that there is no stress-limiting effect at this temperature condition. Further more, since the side surface of Si fin is of (110)/<110> direction and the top surface is of (100)/<110> direction, their oxide growth rate are supposed to be different by 1.4 times. Consider the sidewall oxide growth (34.4 nm) is nearly 1.4 times more than that of top surface (23.2 nm)  we can conclude that difference is mainly due to different crystal orientation on surfaces [15][16].

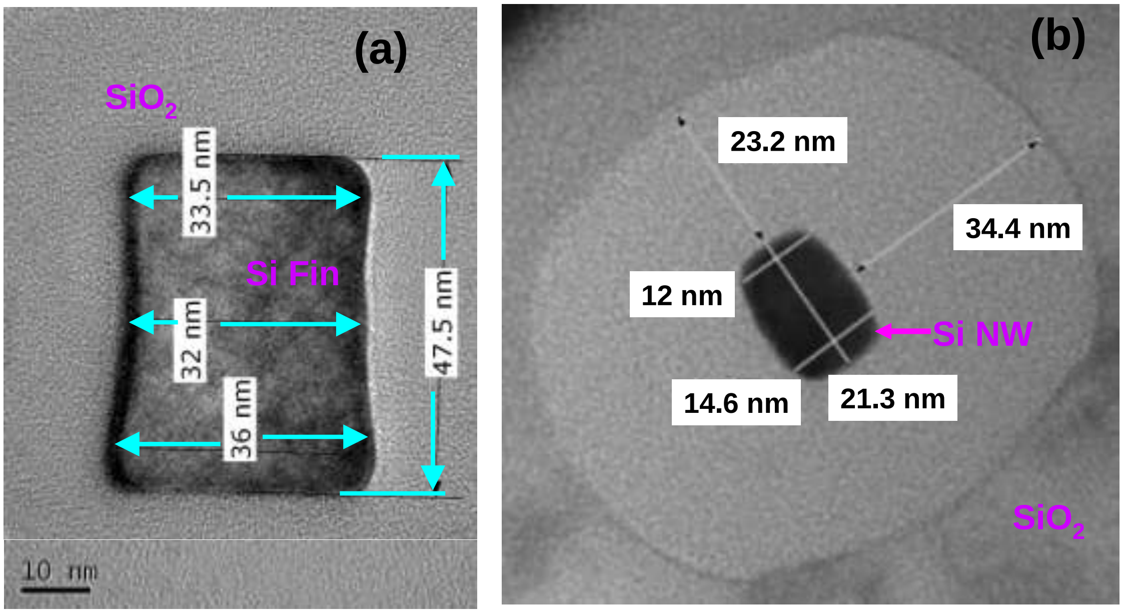


**Fig 4-9**: TEM cross section image of (a) – initial fin before oxidation (b) – after 80 min dry oxidation at 975°C.

**Fig. 4-10** shows the TEM cross section image of a starting fin with the same dimension but has undergone oxidation in 875°C for 1.6, 6 and 15 hrs respectively. It can be seen that the oxide thickness on the sidewall increased by 16.7 nm in the first 1.6 hrs, by 12.2 nm in the next 4.6 hours and by only 3.1 nm in the last 9 hrs. This shows a significant slow down in oxidation rate, resulting in stress-limiting of oxidation. Also, comparing the oxide thickness on top surface and the sidewall, the difference is about 1.6 times, this suggests that the initial tensile stress is playing a role in increasing the oxidation rate on the sidewalls.

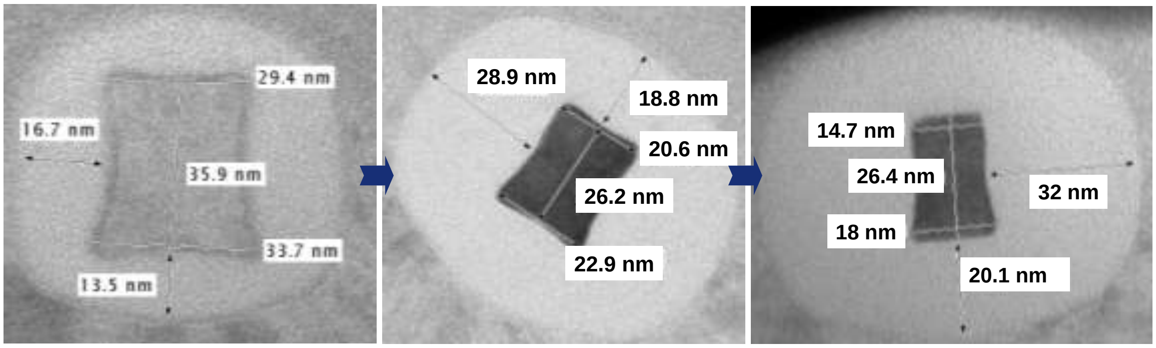


**Fig 4-10**: TEM cross section image of NW channel after 1.6, 6 and 15 hrs oxidation at 875°C respectively

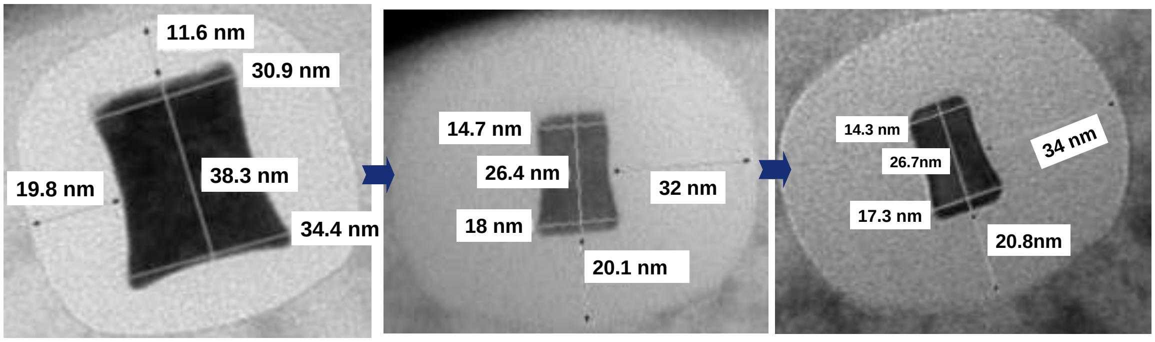


**Fig 4-11**: TEM cross section image of NW channel after 1.7, 13.1 and 22 hrs oxidation at 850°C respectively

**Fig. 4-11** shows the TEM cross section image of a starting fin with the same dimension but has undergone oxidation in 850°C for 1.7, 13.1 and 22 hrs respectively. The oxide thickness on the sidewall increased by 19.8 nm in the first 1.7 hrs, by 12.2 nm in the next 11.4 hours and by only 2 nm in the last 8.9 hrs. Both experiments show strong evidence for stress-limited oxidation. For the 875°C specimens, the remaining Si core size may further reduce for longer oxidation time according to **Fig. 4-5**. However, for the 850°C specimens, we can almost safely conclude that the oxidation is limited to a rate almost zero and the limiting size of Si core and the limiting oxide thickness fits the data shown in **Fig. 4-5** well.

We have experimentally shown that the stress-limiting oxidation can take place on a Si fin of square initial cross section at a temperature as high as 900°C. The initial shape of the fin cross section was found to play a pivotal role in the stress-limiting effects. Our experimental data also shows that the limiting Si core size can be much larger than previously reported (2-10 nm). The understanding of the stress-limiting phenomenon under different conditions is very important for future nanowire process design and control of their uniformity.

## 4.4 Gate stack Formation, Implant and Metallization

Thermally gown oxide was used as gate dielectric for its better interface quality. The devices designed have four gate lengths: 0.2, 0.35, 0.6 and 0.85 µm, for capacitance and mobility study. As shown in **Fig. 4-2**, the gate is not overlapping the S&D region so that the overlap capacitance is reduced. Poly-Silicon is chosen for gate material and was implanted together with the Source and Drain region. Due to the undoped nature of the NW devices, such design results in near 0 V threshold voltages for both N and P-type device.

Due to the small device geometry, alignment accuracy is one of the major challenges for gate stack processes. Misalignment can be detrimental in this fabrication process since it would cause much difference in series resistance and parasitic capacitance in the extremely narrow nanowire. We have carried out extensive lithography trials to ensure the misalignment of the two layers is controlled below 25 nm.

For simplicity of process, the design uses Al/Cu/Si alloy single layer metal for metallization. **Fig. 4-12** (a) shows a SEM image of a NW device after gate stack formation and **Fig. 4-12** (b) shows the SEM image of an inverter circuit after metal etch.

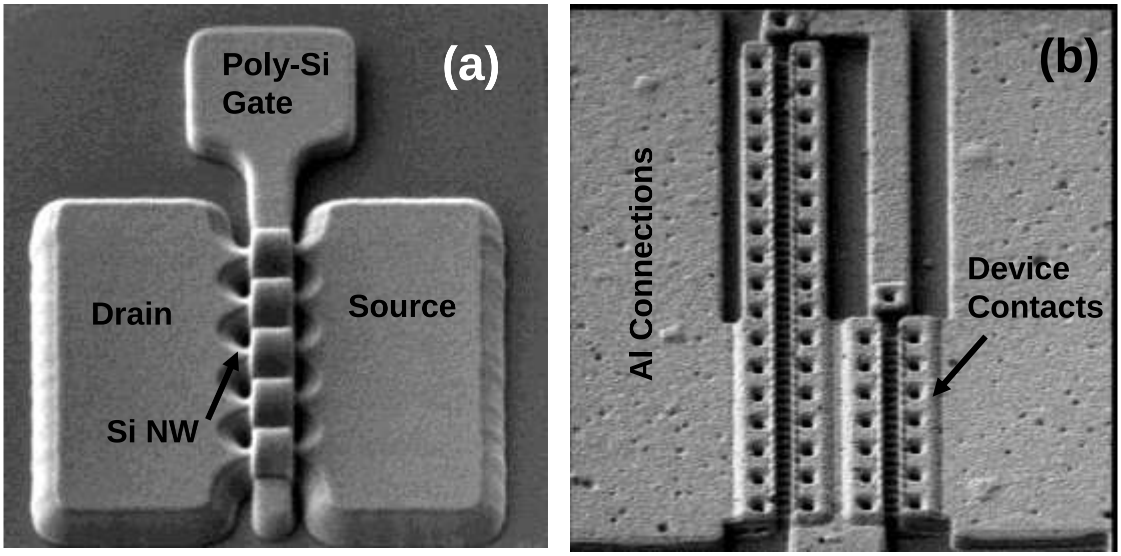


**Fig. 4-12**: (a) - NW device after gate stack formation showing the shape of the poly-Si gate and gate contact (b) - SEM image of an inverter circuit of multi-wire NW devices after metal etch

### *4.5 Poly-Si Stringer Effects and Its Elimination*

Due to the 3D nature of the NW device, its S&D region are isolated islands on B.O.X. layer, creating a topology of 500~700 Å. After blanket deposition of poly-Si and anisotropic gate etch by plasma, a layer of poly-Si residual would remain surrounding the S&D along the sidewall, forming the poly-Si stringer (as shown in **Fig. 4-13**). In most cases, the stringer is still connected to the gate and only separated from the S&D region by a thin gate dielectric layer. This stringer is harmful in many ways:

(1) The stringer increases $C_{gg}$ that results in significant increase in R-C delay in circuit applications

(2) Because the sidewall area of the S&D region is very large comparing with nanowire channel area, the stringer increases gate capacitance dramatically, leading to difficulty in characterizing the channel capacitance.

(3) By increasing the gate area, it also increases the current leakage from gate to drain, leading to unreliable current measurement in CBCM and noise (as described in Chapter 3).

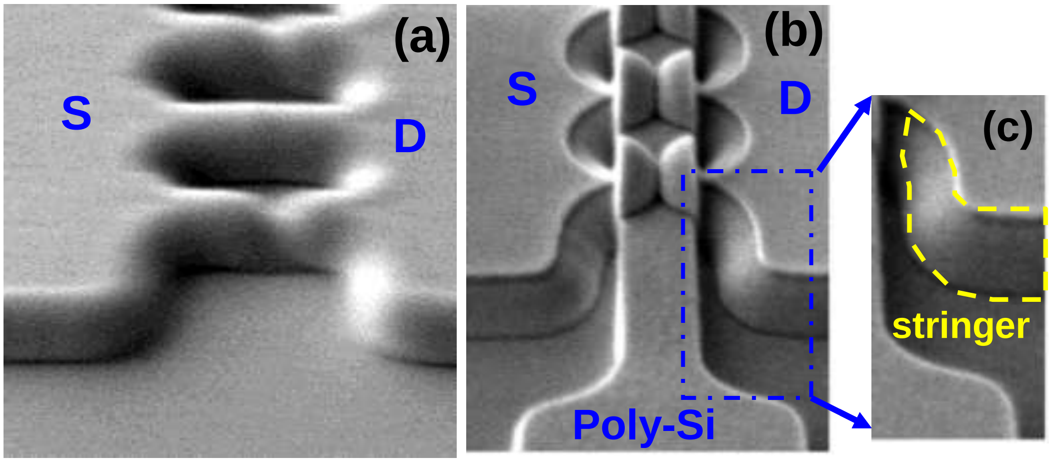


**Fig. 4-13**: The NW device structure after (a) – poly-Si deposition (b) – poly-Si etch showing the stringer surrounding S&D sidewall and connected to the gate (c) – zoomed in image of the stringer (circled with dashed line) formed around the side wall of the source/drain contact region.

To address the problem of stringer, the best way is to eliminate the topology as far as possible. We have developed a modified fabrication process, to fill HDP oxide and planarize the surface using chemical-mechanical polishing (CMP) before poly-Si deposition and etch. To open the window to expose NW channels for gate deposition, an additional mask layer, the reverse gate, is introduced. The reverse gate is patterned on a 150 Å SiN hard mask, which was deposited after CMP to prevent HDP oxide loss during the wet etch process to remove the oxide grown on fin during the conversion of the fin t6o nanowire. Another advantage of this process is that the oxidation of the fin for NW formation can now be down only within the reverse gate, with rest of the device structure being protected by HDP oxide. This prevents the S&D region being thinned down during the NW formation oxidation and will prevent the severe increase in the S/D series resistance of the device. The mask layout with reverse gate layer is shown in **Fig. 4-14**.

The reverse gate length is 100 nm shorter than the gate length (50 nm on each side) by design for all gate lengths. It is important to note that with local fin formation and gate oxidation taking place only within the reverse gate window. With this setup, the channel of the device forms also only within the reverse gate window. Therefore the physical gate length of the device is decided by the length of the reverse gate instead of that defined on the gate mask layer as in the conventional MOS devices. Moreover, the width of gate depends on the total circumferences of the nanowire channels instead of “gate width” defined by poly region in conventional CMOS process. **Fig. 4-15** (a) shows the SEM image of a multiple fin device before CMP. **Fig. 4-15** (b) shows the same device after CMP, reverse gate etching, and local oxidation. It can be seen that the topology of the device is no longer visible in the micrograph. **Fig. 4-15 (c)** shows the device image after 1300 Å poly-Si gate deposition and etching, showing no stringer left. The dent in the middle of the gate is due to poly-Si filling into the reverse gate opening. Thus the critical dimension that determines the physical gate length of the device is the length of the reverse gate instead of the gate.

**Table 4-1** enlists the complete flow of the modified stringer-free uniform cross section silicon nanowire fabrication process. All major steps such as lithography, etching and implant are listed with equipment and process details. We have found that using this method, the parasitic capacitance of the DUT can be reduced from14 fF to less than 1 fF. This greatly reduced the logic gate delay of NW devices and also allows more accurate characterization of the intrinsic gate capacitance.

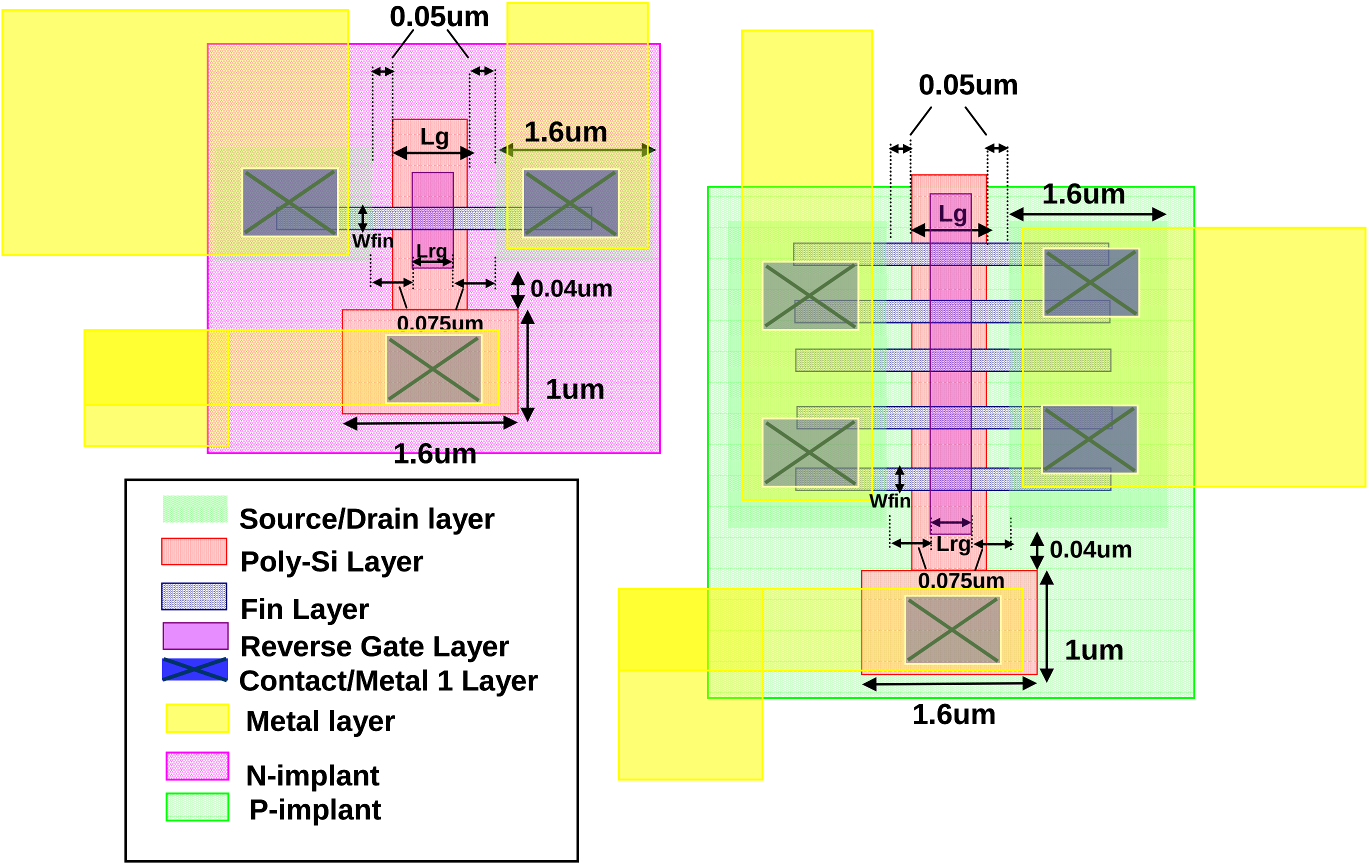


**Fig. 4-14**: Final mask layout design schematic with fin layer and reverse gate layer for both N and P type single and multiple finger NW devices.

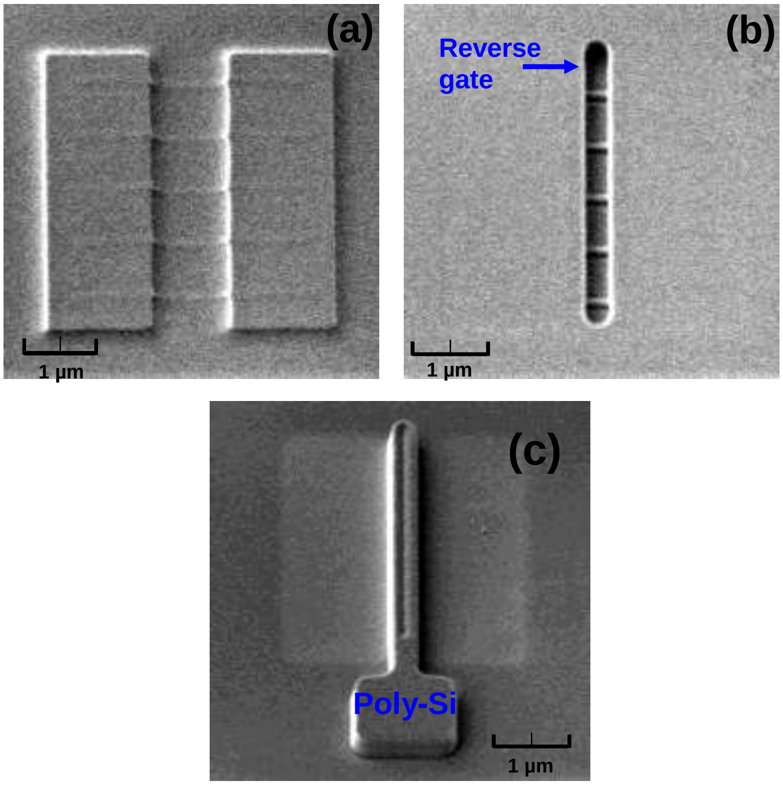


**Fig. 4-15**: (a) – SEM image of a multi-finger NW device before CMP (b) – the same device after CMP, reverse gate etch and local oxidation process showing the successful elimination of topology (c) the same device after poly-Si deposition and etching, showing no stringer.

| Process Step | Equipment | Description |
|---|---|---|
| Wafer preparation | SEMCO furnace<br>FEOL wet bench | SOI layer is thinned down from 70 nm to 50 nm by dry oxidation<br>Remove the oxide using HF solution |
| Hard mask deposition | AMAT PECVD | Si3N4 with 50 Å thermally grown liner oxide are deposited on top of Si for hard mask |
| Lithography - Fin Layer | 248 nm Scanner system | Fin layer patterned |
| Resist trimming | STS Etcher | trim the width of the fin to 50 nm |
| Etch hard mask | P5000 Plasma Etch | Transfer Fin pattern to hard mask |
| Lithography - Source/Drain contact region | 248 nm Scanner system | Source/drain contact region patterned |
| Etch Fin and Source/Drain | P5000 Plasma Etch | Transfer Fin and Source/Drain contact region patterns to Silicon |
| Strip hard mask layer | FEOL wet bentch | Strip off the Si3N4 layer on top of fin using hot H3P04 solution |
| Oxidation - liner oxide | SEMCO furnace | grown 50 Å thermal oxide as liner |
| HDP deposition and condensation | AMAT PECVD<br>SEMCO furnace | Deposit 2500 Å HDP and condense in furnace at 900ºC |
| Chemical mechanical polishing | Varian CMP | polish to remove ~2000 Å HDP to reduce wafer surface topology |
| Hard mask deposition | AMAT PECVD | Si3N4 hard mask deposited for reverse gate opening |
| Lithography - Reverse gate | 248 nm Scanner system | Reverse gate region patterned |
| Etch reverse gate | TEL oxide etcher<br>FEOL wet bench | Etch reverse gate region, stop before reaching fin<br>Remove residual oxide on fin sidewall by wet etching in HF solution |
| Nanowire formation | SEMCO furnace<br>FEOL wet bench | Use dry oxidation to consume the silicon fin (by a combination of high and low temperature)<br>Remove the grown oxidation in HF solution, leaving only a thin Si core. |
| Gate dielectric layer | SEMCO furnace | Thermally grown oxide on si nanowire as gate dielectric |
| Poly Si deposition | AMAT PECVD | Deposit 1300 Å poly-Si as gate electrode material |
| Lithography - Poly-Si layer | 248 nm Scanner system | gate layer patterned |
| Etch poly gate | P5000 Plasma Etch | Transfer gate pattern to poly-Si |
| Hard mask strip | FEOL wet bentch | Strip off the Si3N4 layer residing out side poly-Si region using hot H3P04 solution |
| Lithography - P-type dopant implant | 248 nm Scanner system | P-implant layer patterned |
| Implant | Varian Ion Implanter | Implant BF2 to gate and source/drain contact region for P devices |
| Lithography - N-type dopant implant | 248 nm Scanner system | N-implant layer patterned |
| Implant | Varian Ion Implanter | Implant phosphorus to gate and source/drain contact region for N devices |
| Dopant activation aneal | Anealing furnace | Aneal after implant using high temperature spike aneal |
| Pre-metal dielectric deposition | AMAT PECVD | Deposit 4000 Å oxide as PMD |
| Lithography - contact | 248 nm Scanner system | contact layer patterned |
| Etch contact | TEL oxide etcher | open contact region through PMD and HDP |
| Metal deposit | Centrura PVD | Deposit TaN/AlSiCu/TaN |
| Lithography - metal | 248 nm Scanner system | metal ayer patterned |
| Etch metal | Centrura Metal etcher | Etch metal to form connection lines and contact pads |
| Post metal clean | FEOL wet bench | Clean metal debris and polymer |
| Alloying | Sintering furnace | Heat treatment of metal for alloying |

**Table 4-1:** Complete process flow for stringer-free, uniform cross section NW device fabrication on CMOS platform

## *4.6 Chapter Summary and Conclusion*

In this chapter, we introduced the fabrication process of NW transistor based on the traditional CMOS technology platform. Key technology issues such as the lithography resolution constraint, the stress-limited oxidation for NW formation has been discussed and a process flow has been described for the elimination of stringer. These challenges were met successfully and their solutions are presented in detail. Also, basic theories related to stress-limiting oxidation were elaborated with modeling using 2D simulation. The inadequacy of SUPREM4 default model in predicting the self-limiting oxidation was also discussed.

Using the process described, we are able to fabricate stringer-free NW device with various sub-micron channel lengths and wire diameter as small as 10 nm. Logic circuits such as inverters, ring oscillators and SRAMs, as well as test circuit for charge-based capacitance measurement, were also fabricated for study and characterization of the NW devices.

## *Reference for Chapter 4*

[1] Y. Xia, P. Yang, Y. Sun, Y. Wu, B. Mayers, B. Gates, Y. Yin, F. Kim, and H. Yan, “One-dimensional nanostructures: Synthesis, characterization, and applications,” *Adv. Mater.*, vol. 15, no. 5, pp. 353–389, Mar. 2003.

[2] G. Li, N. Xi, H. Chen, A. Saeed, and M. Yu, “Assembly of nanostructure using AFM based nanomanipulation system,” in *Proc. IEEE Int. Conf. Robot. Autom.*, 2004, vol. 1, pp. 428–433.

[3] Y. Huang and C. M. Lieber, “Integrated nanoscale electronics and optoelectronics: Exploring nanoscale science and technology through semiconductor nanowires,” *Pure Appl. Chem.*, vol. 76, no. 12, pp. 2051–2068, 2004.

[4] H. Ye, Z. Gu, and D. H. Gracias, “Integrating nanowires with substrates using directed assembly and nanoscale soldering,” *IEEE Trans. Nanotechnol.*, vol. 5, no. 1, pp. 62–66, Jan. 2006.

[5] N. A. Melosh, A. Boukai, F. Diana, B. Gerardot, A. Badalato, P. M. Petroff, and J. R. Heath, “Ultrahigh-density nanowire lattices and circuits,” Science, vol. 300, no. 5616, pp. 112–115, Apr. 2003.

[6] D. W. Wang, B. A. Sheriff, and J. R. Heath, “Complimentary symmetry silicon nanowire logic: Power-efficient inverters with gain,” Small, vol. 2, no. 10, pp. 1153–1158, 2006.

[7] R. He, D. Gao, R. Fan, R. Hochbaum, C. Carraro, R. Maboudian, and P. Yang, “Si nanowire bridges in microtrenches: Integration of growth into device fabrication,” Adv. Mater., vol. 17, no. 17, pp. 2098–2102, Apr. 2005.

[8] W. Lu and C.M. Lieber, "Semiconductor nanowires," J. Phys. D: Appl. Phys. 39, R387-R406, 2006.

[9] W. Lu, P. Xie and C. M. Lieber, “Nanowire transistor performance limits and applications”, *IEEE Tran. Elect. Dev.* Vol. 55, no. 11, pp. 2859-2876, 2008

[10] T. Tezuka, E. Toyoda, S. Nakaharai, T. Irisawa, N. Hirashita, Y. Moriyama, N. Sugiyama, N. Taoka, Y. Yamashita, O. Kiso, M. Harada, T. Yamamoto, and S. Takagi, “Observation of mobility enhancement in strained Si and SiGe tri-gate MOSFETs with multi-nanowire channels trimmed by hydrogen thermal etching,” in IEDM Tech. Dig., 2007, pp. 887–890.

[11] S. D. Suk, S.-Y. Lee, S.-M. Kim, E.-J. Yoon, M.-S. Kim, M. Li, C.W. Oh, K. H. Yeo, S. H. Kim, D.-S. Shin, K.-H. Lee, H. S. Park, J. Nam, H. C. J. Park, J.-B. Park, D.-W. Kim, D. Park, and B.-I. Ryu, “High performance 5 nm radius Twin Silicon Nanowire MOSFET (TSNWFET): Fabrication on bulk Si wafer, characteristics, and reliability,” in IEDM Tech. Dig., 2005, pp. 552–555.

[12] H. I. Liu, D. K. Biegelsen, N. M. Johnson, F. A. Ponce, and R. F. W. Pease, “Self-limiting oxidation of Si nanowires,” J. Vac. Sci. Technol. B, Microelectron. Process. Phenom., vol. 11, no. 6, pp. 2532– 2537, Nov./Dec. 1993.

[13] W. M. Moreau, Semiconductor Lithography, Plenum Publishing Co., New York, 1988.

[14] H. I. Liu, D. K. Biegelsen, R. F. W. Pease, N. M. Johnson and F. A. Ponce, "Self-limiting oxidation of Si Nanowires", *J. Vac. Sci. Technol. B* vol.11, no.6, pp. 2532-2537, 1993.

[15] E. H. Nicollian and J. R. Brew, *MOS Physics and Technology,* New York, Willey, pp. 681-689, 1982.

[16] C. Gonzalez and J. P. McVittle, “A study of trenched capacitor structures”, *IEEE Electron Device Lett.,* vol. 6, no. 5, pp. 215-218, 1985.

# Chapter 5

# Sub-Femtofarad Scale Capacitance Measurement Using CBCM

## *5.1 Introduction*

There has been much excitement being generated in recent years about semiconductor nanowires (NWs) for their potential performance to exceed that of conventional devices substantially [1][2]. However, on experimental side, there is ambiguity in mobility values. The main issue concerning these devices has been that the lack of direct measurements of single NW capacitance, leading to accurate extraction of channel mobility values.

The CV-Meter type of equipment available for capacitance measurement either measure the current of an ac voltage source at high frequencies or determine the charge induced by a small voltage step (as in a quasistatic CV-Meter). The accuracy of such measurements suffers when the capacitance under test is below the femto-farad level due to the poor signal to noise ratio. Another method uses an S-parameter test set to measure the capacitance through the measurement of the admittance $Y_{11,}$ in which high accuracy can only be achieved when the measured impedance matches the internal impedance of 50 **Ω**. In all cases, the capacitance being tested has to be connected directly to the terminals of the instrument which increase the background capacitance further due to the capacitance of the connection. In order to avoid measurement errors coming from the test connection and environment, costly shielding techniques have to be applied. Also, problems arise from parasitic capacitances on the chip itself, e.g., pads and wiring, which can be as high as the capacitance being measured. Therefore, for fF range capacitance measurement using the above-mentioned instruments, complicated de-embedding procedures are needed. Tu et al. [3] reported the carefully measured C-V and mobility of a single channel of grown Ge NW FET, but only at cryogenic temperatures. As most of the conventional C-V measurements, their method still suffers from high background capacitance and high noise especially at room temperature.

Guided by results and findings of the simulation study on CBCM in Chapter 3, we carry out experimental CBCM measurement on fabricated test keys with nanowire FET as device-under-test (DUT) and report the results in this chapter. We first introduce the CBCM test key design and fabrication, followed by a demonstration of an innovative approach to CBCM [4] to measure C-V and I-V characteristics from the same Silicon NW test structure. The test setup will be benchmarked against conventional C-V for validation using larger capacitance values. Finally, the measurement results on single channel NW devices will be presented.

## 5.2 CBCM Test Key Design and Fabrication

Based on the extensive simulation studies delineated in Chapter 3 and its conclusion, we have chosen the charge-injection free scheme proposed by Chang et al [5] with the provision of measurement of the I-V characteristics from the same characterized device. Its major advantage is the total elimination of the error introduced by mismatch and process variation which is almost inevitable in nano-scale fabrication. Moreover, it addresses the charge-injection problem by having the DUT physically connected to the test key in both measurements [6]. Therefore, it is the most practical method for the first trial application of CBCM to the sub-femto farad capacitance measurement. Shown in **Fig.5-1** is the schematic of the test key circuit with its 6 input/output terminals. The source and drain terminals (shown as *S* and *D*) of the device under test (DUT) were deliberately separated for measuring its I-V characteristic.

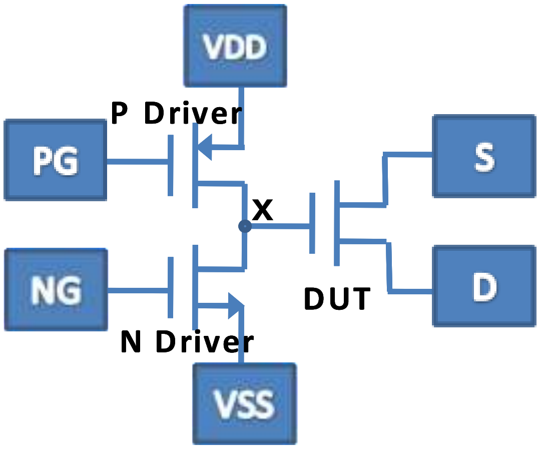


**Fig. 5-1**: CBCM Test key designed to measure C-V and I-V on the same NW DUT.

Test keys were designed for both N and P-type NW DUT with four gate lengths: 0.2 µm, 0.45 µm, 0.6 µm, and 0.85 µm. Besides single wire DUTs, multiple finger DUT with 5, 20 and 100 wires connected in parallel were also included, each having two pitches, 0.4 µm and 0.6 µm, for the study of parasitic and pitch effects in NW devices. There are a total of 112 test keys fabricated.

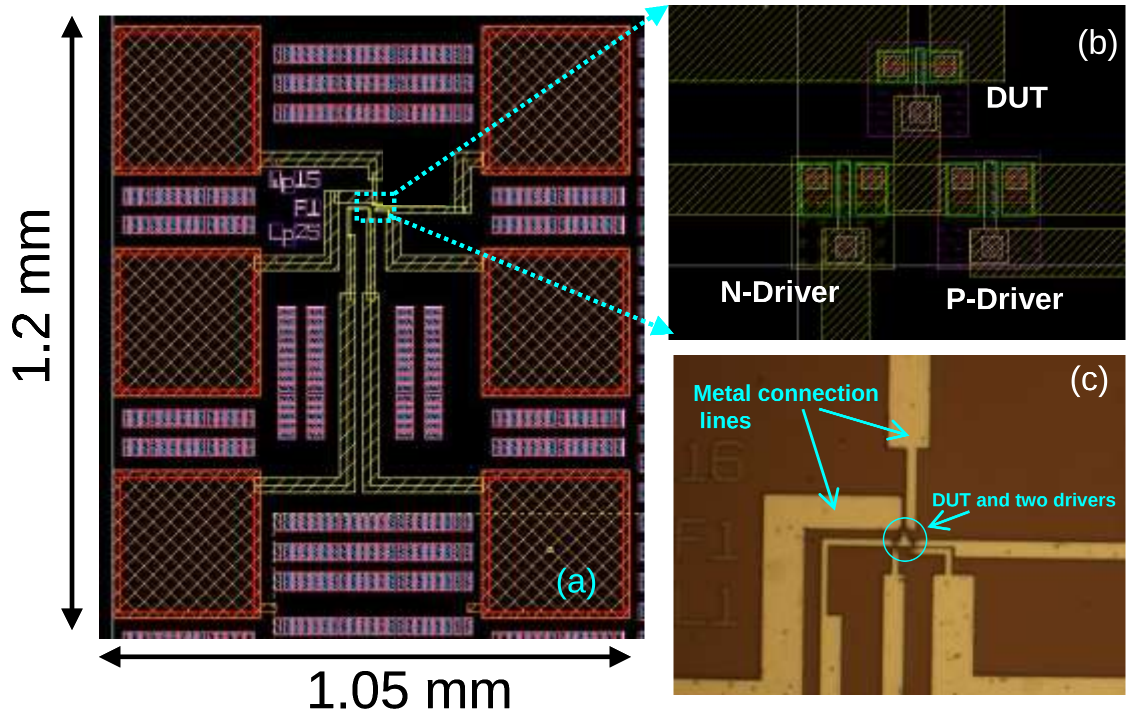


**Fig. 5-2**: (a) – complete layout of a CBCM test key (b) – zoomed in image of the devices in the test key (c) – a micro image of the test key fabricated on-chip showing the DUT, the drivers and the metal connections to testing pads.

**Fig. 5-2** (a) shows an example of a test key circuit layout with input/output pads. As we have shown in Chapter 3, to reduce the errors introduced in CBCM measurement, the estimated DUT capacitance must be at least half of the parasitic capacitance at the drain of pseudo-inverter (denoted as node "X" in Fig. 5-1). Therefore, the drain of the drivers and the gate contact of DUT are placed as close to each other as possible to reduce the parasitic capacitance introduced by metal lines as shown in **Fig. 5-2** (b). **Fig. 5-2** (c) shows a micro image of the test key fabricated on-chip.

Both the DUT and driver devices are NW transistors except that the drivers have a larger NW channel diameter so that it can charge/discharge DUT and parasitic capacitances effectively. The test keys are fabricated using the process technology described in Chapter 4. **Fig. 5-3** shows some images of the fabricated test key and the

NW channels. We have used 10nm thick deposited oxide as gate dielectric for this experiment. The thickness of dielectric is on the higher side considering that these structures have potential of scaling the channel lengths to tens of nanometers, there is obviously an advantage for this experiment as it reduces the gate capacitance and takes the CBCM technique of the capacitance measurements towards its lower limits or towards highest resolution.

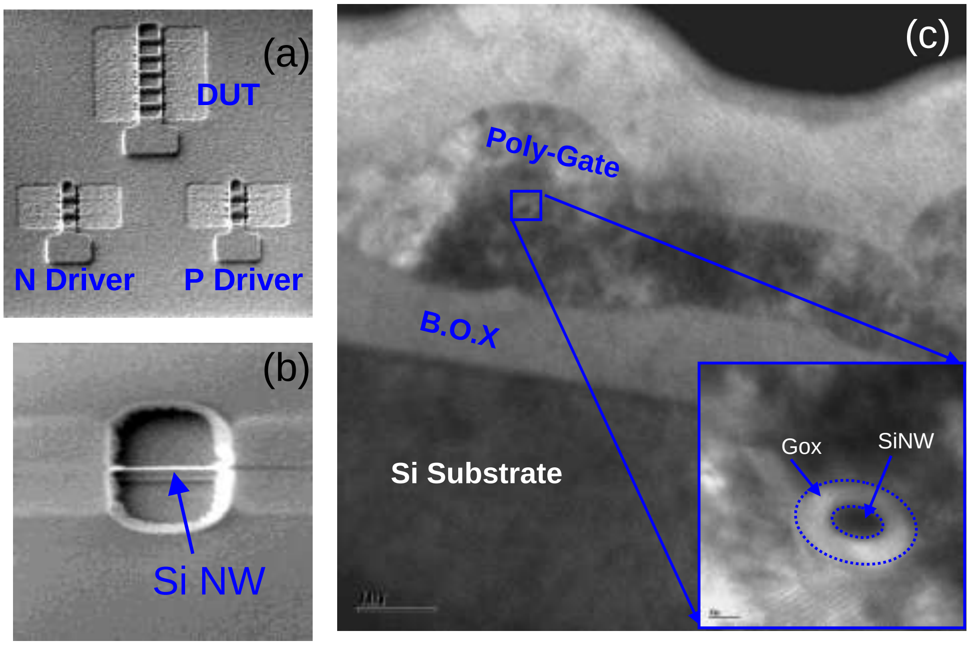


**Fig. 5-3**: (a) – SEM image of a test key set fabricated on-chip with a 5 wire multiple finger DUT (b) – SEM image of a single wire NW DUT after local release (c) – TEM image of the NW cross section after gate formation.

## *5.3 CBCM Test setup and I-V Characterization of DUT and Drivers*

In CBCM measurement, the two charging (or discharging) currents through the terminal *VDD*, $I_{V_{DD}}$ and $I'_{V_{DD}}$ , with constant voltage and non-overlapping pulse applied respectively to the *S* and *D* contact regions, are monitored. As seen in **Fig. 3-2 (b)**, when a constant voltage is applied to the S/D of the DUT, the current through *VDD* charges/discharges the gate-to-drain and gate-to-substrate parasitic capacitance of the pseudo-inverter as well as the DUT channel capacitance. When a third non-overlapping pulses is applied to a S/D of DUT, pre-charging/discharging its channel to 0 or *VDD* level, only the parasitic part is the being charged/discharged. The bias dependent capacitance $C_{gg}$ of the DUT is then derived by differentiating the difference of $I_{V_{DD}}$ and $I^{'}_{V_{DD}}$ with respect to the voltage applied to *VDD* as shown in Eqn. 5-1:

$$C_{DUT} = \frac{d(Q_{V_{DD}} - Q^{'}_{V_{DD}})}{dV_{DD}} = \frac{d(I_{V_{DD}} - I^{'}_{V_{DD}})}{dV_{DD}} \frac{1}{f} \qquad \textit{(5-1)}$$

The pulses applied to the terminals of the test key for N and P-type DUTs are shown in **Fig. 5-4** (a) and (b) respectively. Take the N-type DUT case in **Fig. 5-4** (a) as an example, the non-overlapping pulses PGn and NGn with the same frequency are applied to the gates of the P and N-driver devices respectively. At the same time, signal $\mathrm{SDn}_1$, a constant voltage, is applied to the source and drain terminals of the DUT. During the rising edge of PGn, the P-driver is being turned off, while N-driver is also off. If the charge injection is not taken into account, the voltage at the gate of DUT (shown in **Fig. 5-1** as node X) should remain at $V_{DD}$ level until the N-driver is turned on by the rising edge of NGn, which discharges the capacitances connected to node X to ground. After that, the falling edge of NGn turns the N-driver off while the voltage at node X remains at ground. When the P-driver is being turned back on by the falling edge of the PGn, the voltage at node X will again rise to $V_{DD}$, charging the capacitances connected to it again. This capacitance being charged and discharged includes not only the gate capacitance of the DUT but also the parasitic capacitance of the driver devices. In order to decouple the effect of the parasitic, the source and drain terminal would need to be pre-charged to either $V_{DD}$ or the ground level before being charged/discharged the pulses PGn and NGn. Therefore, in the second step of the measurement, instead of a constant voltage, a third non-overlapping pulse $SDn_2$ is applied to the source and drain terminals of DUT. Its falling edge occurs after the rising edge of PGn, which allows the gate capacitance of the DUT to be discharged to ground before the rising edge of NGn. On the other hand, its rising edge occurs after the falling edge of NGn, allowing the gate capacitance of DUT to be charged back to $V_{DD}$ before the falling edge of PGn. As a result, only parasitic capacitance connected to the node X is being charged/discharged in the second measurement step. Applying the measurement result to **Eqn. 5-1**, the gate capacitance of DUT can be obtained. For P-type DUT device, the polarity of the pulses are reversed, but the general principle described above remains the same.

Voltage applied to $V_{DD}$ needs to be swept across the measurement range in order to derive voltage dependent capacitance. Also, in order to enhance the accuracy of measurement of $C_{gg}$ of the P-type DUT in inversion region, we applied a negative voltage to $V_{SS}$ and 0V to $V_{DD}$ instead and appropriately altered the pulse levels to gate of the drivers as shown in **Fig. 5-4** (b).

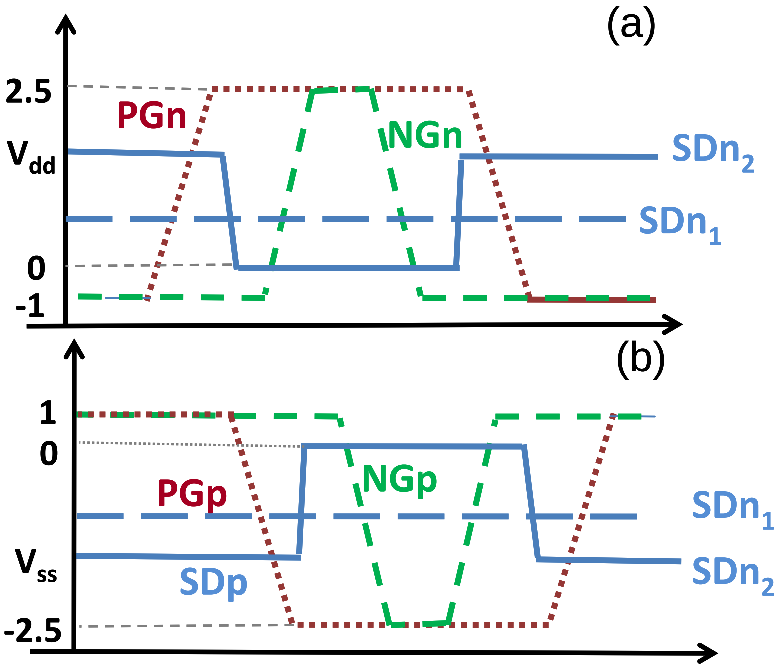
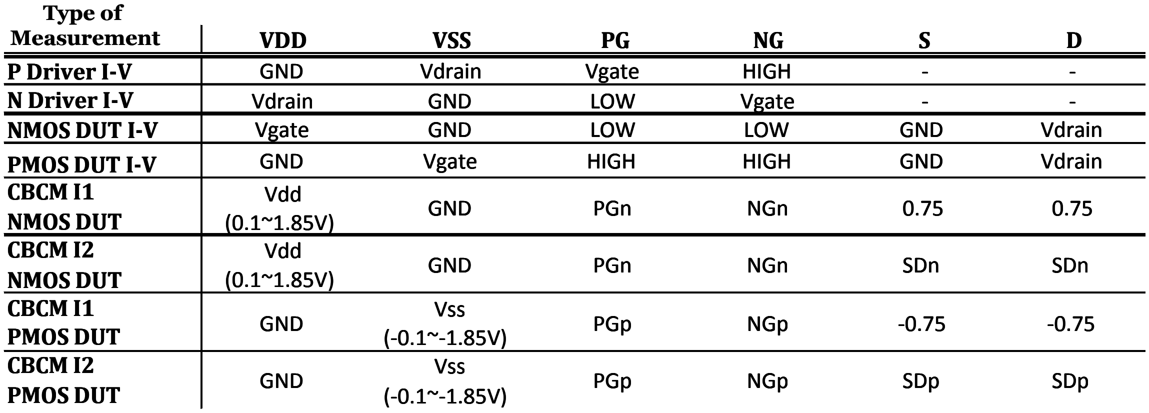


**Fig. 5-4**: Pulse inputs to PG, NG, S and D terminals (Fig. 5-1) (a) for N-type and (b) for P-type DUT C-V measurements.

Fabrication of nanowire devices brings in inevitable device variability due to the non-uniform conditions across dies and wafers. In a NW device of sub-10 nm radius, a few nanometer of difference could result in large deviation in device characteristics such as current, capacitance, and threshold voltage [7]. Therefore, in order to avoid distortion due to variability and obtain sensible mobility result, it is essential that the drivers and DUT currents and series resistance can also be characterized using the same test key. This step not only prevents the leakage current from distorting the CBCM result, it also avoids the ambiguity due to process variability in NW devices. Forcing the P driver on, the N drive I-V characteristic can be measured by applying a gate voltage to *NG* and drain voltage to $V_{DD}$. In the same way, P driver I-V can be measured by forcing N driver on, and applying the gate and drain voltage to *PG* and $V_{SS}$ while grounding the $V_{DD}$. Similarly, the DUT I-V characteristic can be measured by having one of the drivers on and the other off. Gate voltage and Drain voltage can be supplied to $V_{DD}/V_{SS}$ and *S/D*. **Table 5-1** enlists the stimulus applied to different terminals for different measurements.



| Type of Measurement | VDD | VSS | PG | NG | S | D |
|---|---|---|---|---|---|---|
| P Driver I-V | GND | Vdrain | Vgate | HIGH | - | - |
| N Driver I-V | Vdrain | GND | LOW | Vgate | - | - |
| NMOS DUT I-V | Vgate | GND | LOW | LOW | GND | Vdrain |
| PMOS DUT I-V | GND | Vgate | HIGH | HIGH | GND | Vdrain |
| CBCM I1<br>NMOS DUT | Vdd<br>(0.1~1.85V) | GND | PGn | NGn | 0.75 | 0.75 |
| CBCM I2<br>NMOS DUT | Vdd<br>(0.1~1.85V) | GND | PGn | NGn | SDn | SDn |
| CBCM I1<br>PMOS DUT | GND | Vss<br>(-0.1~-1.85V) | PGp | NGp | -0.75 | -0.75 |
| CBCM I2<br>PMOS DUT | GND | Vss<br>(-0.1~-1.85V) | PGp | NGp | SDp | SDp |

**Table 5-1:** Stimuli applied at the terminals of the test key for measuring the I-V characteristics of the N and P drivers and the DUT, and, the C-V measurement of N and P-type SNW DUTs. The I-V of Driver devices help in deciding the values of LOW and HIGH and the levels of PGn, PGp, NGn, and NGp pulses in the table. The drivers are SNW devices with gate length of 0.35μm and a larger diameter to source/sink sufficient charging/discharging current.

Using the setup described in **Table 5-1**, we carried out I-V measurements on both the DUT and the N and P-type drives, as shown in **Fig. 5-5** (a) and (b) respectively. Due to the fully-depleted nature of the NW device, and the use of poly-Si gate, the $V_{th}$ of both N and P-type device were near zero. On the other hand, the driver devices, which has a larger NW channel diameter, shows different $V_{th}$, with $V_{th.Ndriver}$=-0.51 V and $V_{th,\ Pdriver}$=0.1 V.

Taking the threshold voltage into consideration, we adjusted the voltage level of the pulses as well as the voltage applied to the drivers' gates to ensure proper CBCM measurement i.e. the drivers turn fully 'on' and fully 'off' as required by the scheme. For N-type DUT, the range of $V_{DD}$ is made 0.1~1.85 V and for P-type device, it is -1.85~ -0.1 V as enlisted in **Table 5-1**. Although the gate terminal of DUT is being charged and discharged between $V_{DD}$ and $V_{SS}$, to measure both polarity of the C-V, a +/- 0.75 V is applied to the *S* and *D* contact regions while measuring $I_{V_{DD}}$ . Therefore the actual range of

$V_{gsd}$ applied is -0.65~1.1 V and -1.1~0.65 V for N and P-type DUT respectively. To achieve accurate differentiation and smoother C-V curves, $V_{DD}/V_{SS}$ are swept with fine steps of on 10 mV. Also, to prevent the increasing gate leakage current at higher $V_{gs}$ distorting the measurement of charge, the pulse heights of the non-overlapping signals to *PG* and *NG* are varying with respect to the $V_{DD}$ level to maintain a fixed overdrive voltage on the P and N-driver devices.

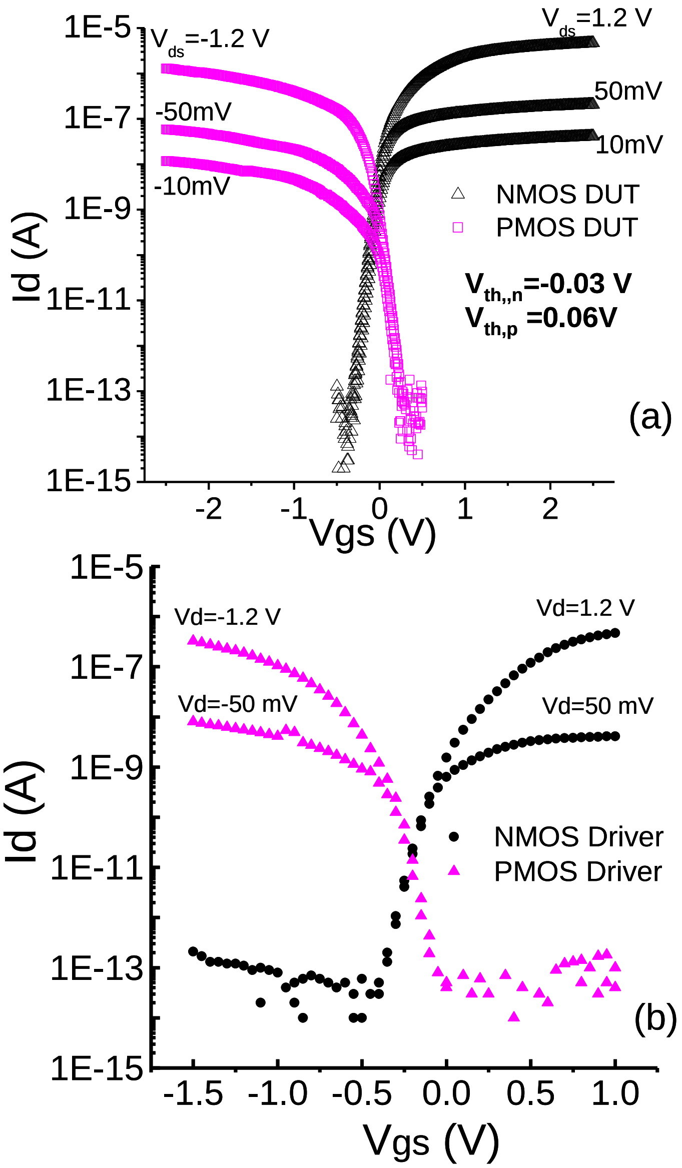


**Fig. 5-5**: (a) – Measured $I_d$ –$V_{gs}$ characteristics for P and N- type SNW DUT under various bias conditions. The threshold voltages of both N and P devices are extracted using Gm (Max) technique. (b) - : Measured $I_d$ –$V_g$ characteristics for P and N-type driver devices. These devices have wider channel diameters and thus different threshold voltages.

## 5.4 Verification of CBCM with LCR Meter Measurement

To establish that the test setup works well, we first benchmarked its measurement results against the conventional C-V meter measurement. Since single channel device capacitance remains beyond the measurement range of conventional C-V measurement, we carried out the verification on a 100-finger NW device which has a gate capacitance in the range of 15~50 fF. Although sophisticated shielding is not available at the time of characterization, care was taken to calibrate the LCR (HP4284) meter used for C-V measurement and minimize the noise. To reduce the random noise in measurements, an average of 16 C-V measurements was used as the measured capacitance is much lower than the instrument's recommended range.

It was found that even after careful calibration, the LCR meter showed a residual capacitance in the range of +/-2fF. With good confidence and repeatability, we were able to measure the gate capacitances of both N and P-type 100-finger NW DUT devices. Their capacitances were then measured using the CBCM setup described in the previous section. As shown in **Fig. 5-6** (a) and (b), the capacitances derived from CBCM technique matches well with that from LCR meter for both N and P-type DUT considering the residual capacitance floor of +/-2fF. As there is no fundamental change of measurement mechanism for much smaller DUT capacitances, we believe this verification result still holds true for sub-fF capacitance measurement using CBCM.

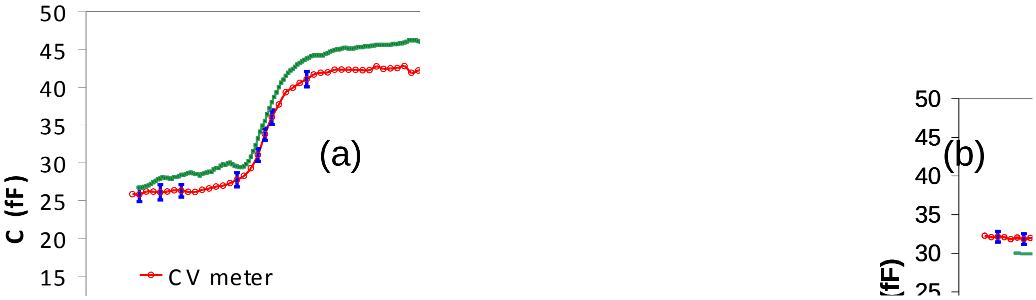


**Fig. 5-6**: Calibration of CBCM technique using precision LCR meter (Agilent 4284). The DUT in this case are 100 finger N (a) and P type (b) SNW devices. The LCR meter capacitances are influenced by the residual capacitance after calibration which is in the range of ±2 fF. The error bars in the C-V meters indicate standard deviation of 16 measurements.

## 5.5 Femto-farad scale capacitance measurement using CBCM

As we have seen in Chapter 3, the charging/discharging current of CBCM is time varying and its magnitude is zero most of the time in a pulse period. Due to the sharp rising/falling of current magnitude, it is necessary to measure it with a higher sampling rate so that the charge can be accurately integrated. Also, the output using current measuring instruments (Agilent 4156C in our case) is the average current which is actually the integrated total charge divided by the whole pulse period. **Fig. 5-7** compares the time-dependent currents $I_{V_{DD}}$ and $I'_{V_{DD}}$ , extracted from TCAD simulation and the average current measured as a function of $V_{gsd}$ applied to the DUT. While there is no constant current flowing through the test keys through out the pulse cycle, most of the charging and discharging of the capacitances happens as sharp peaks during edges of the non-overlapping pulses, as shown in the zoomed in current plots in **Fig. 5-7** (b)-(c). When these charging/discharging currents are being measured by a current meter, the measurements are constant time-dependent currents. **Fig. 5-8** is shows the difference of the two currents measurements as a function of $V_{gsd.}$ With the applied pulse frequency known (1 MHz in this case), the capacitance ($C_{par}$+$C_{DUT}$), $C_{par}$ and $C_{DUT}$ as a function of $V_{gsd}$ can be derived by differentiating these curves.

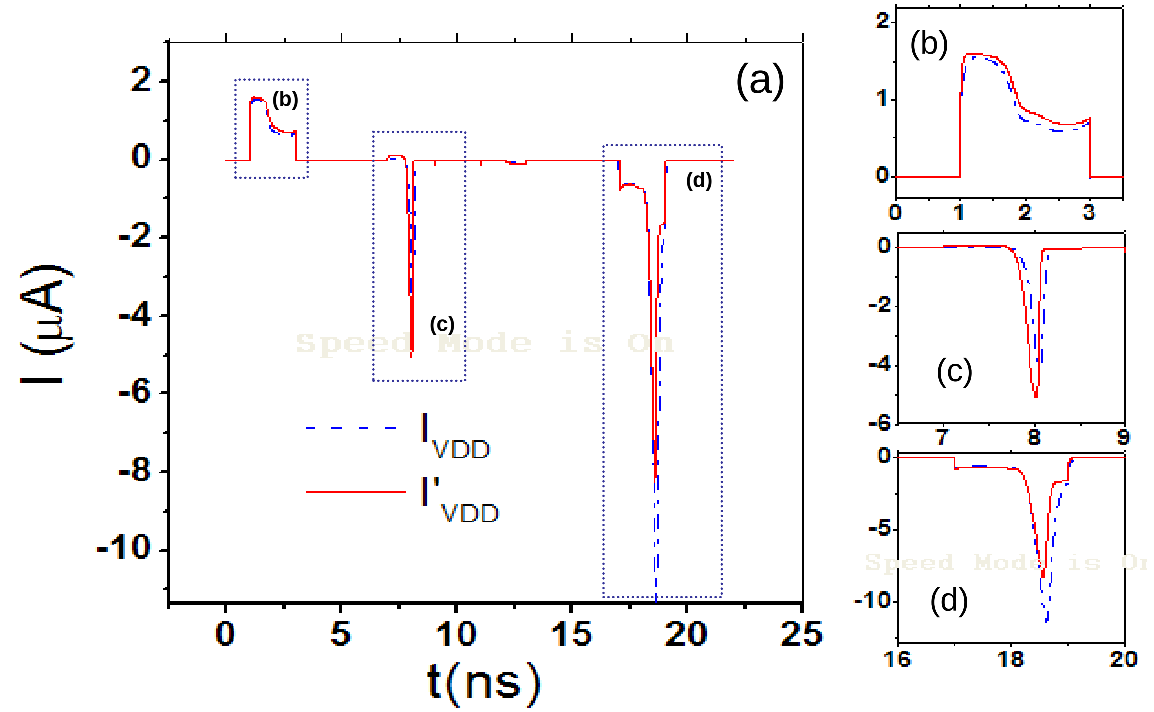


**Fig. 5-7:** (a) $I_{V_{DD}}$ and $I'_{V_{DD}}$ plotted as a function of time extracted by TCAD simulation. The time dependent currents rise/fall sharply during charging discharging phases (b) – (c) show the zoomed plots of the three current peaks

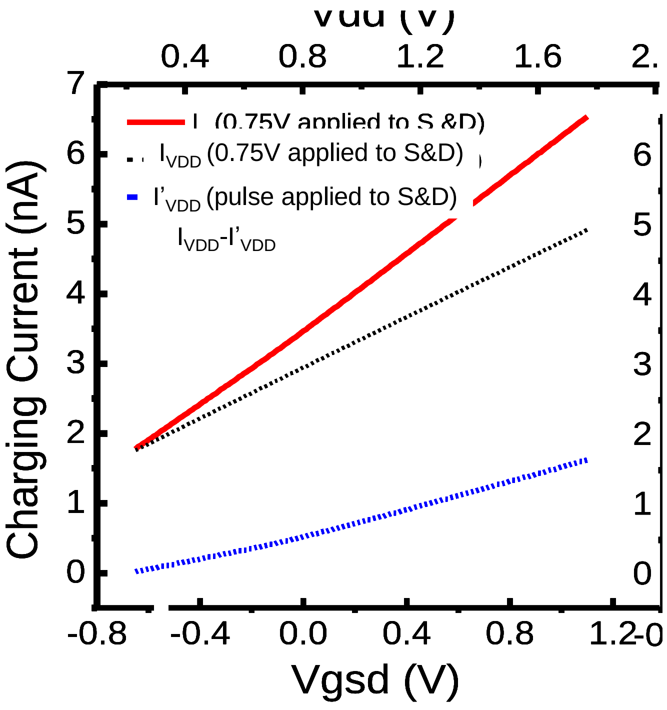


**Fig. 5-8:** Measured average currents of $I_{V_{DD}}$ and $I'_{V_{DD}}$, as well as their difference ( $I_{V_{DD}}$ - $I'_{V_{DD}}$ ) plotted as a function of $V_{gsd}$.

Since capacitance $C_{DUT}$ is derived by differentiation of the difference of the two small currents, any error gets amplified due to loss of significant digits in this process. To deal with this, 16 measurements were taken for each device and outliers, if any, were removed following Peirce's criteria [8] before taking the averages. Noise free numerical derivatives are calculated following Savitzky and Golay's algorithm [9]. **Fig. 5-9** shows the capacitance ($C_{par}$+$C_{DUT}$), $C_{par}$ and $C_{DUT}$ extracted for single finger N and P-type DUT of 0.85 μm long gates.

The $C_{par}$ curves for P and N-type devices show significantly different voltage dependence. This is due to the gate-to-substrate parasitic capacitance. In case of N-type device, the substrate remains in depletion/inversion and shows constant (low) capacitance for positive voltage applied to DUT gate while it changes its state from accumulation to inversion when negative voltage range is applied to the DUT gate for PMOS. Also, the P-type DUT device has lower minimum capacitance than N-type DUT device perhaps on account of variation in dimension from device to device and difference in doping levels in the gate electrode.

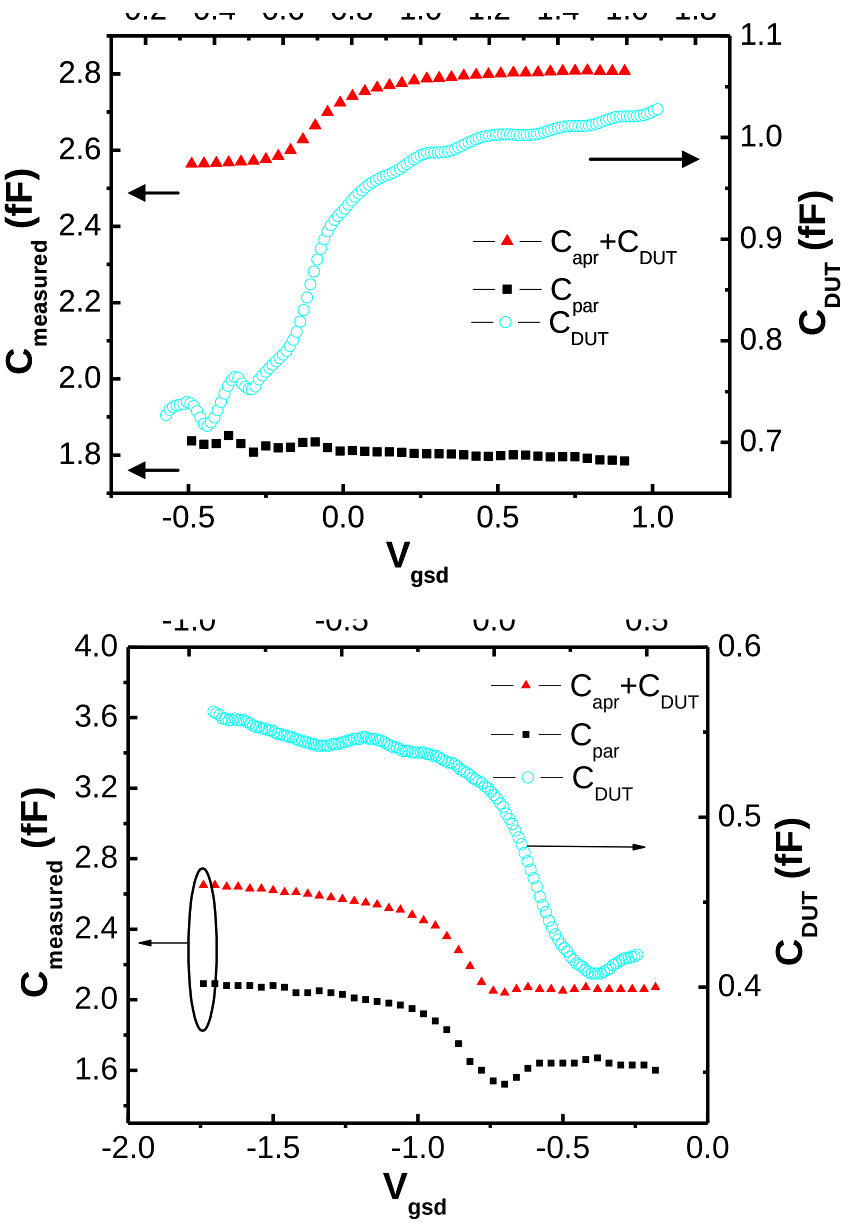


**Fig. 5-9**: CBCM measurement of $C_{par}$, ($C_{par}$+$C_{DUT}$) and the very low $C_{DUT}$ of ~1fF for (a) N-type (b) P-type single finger SNW device of 0.85 μm gate length.

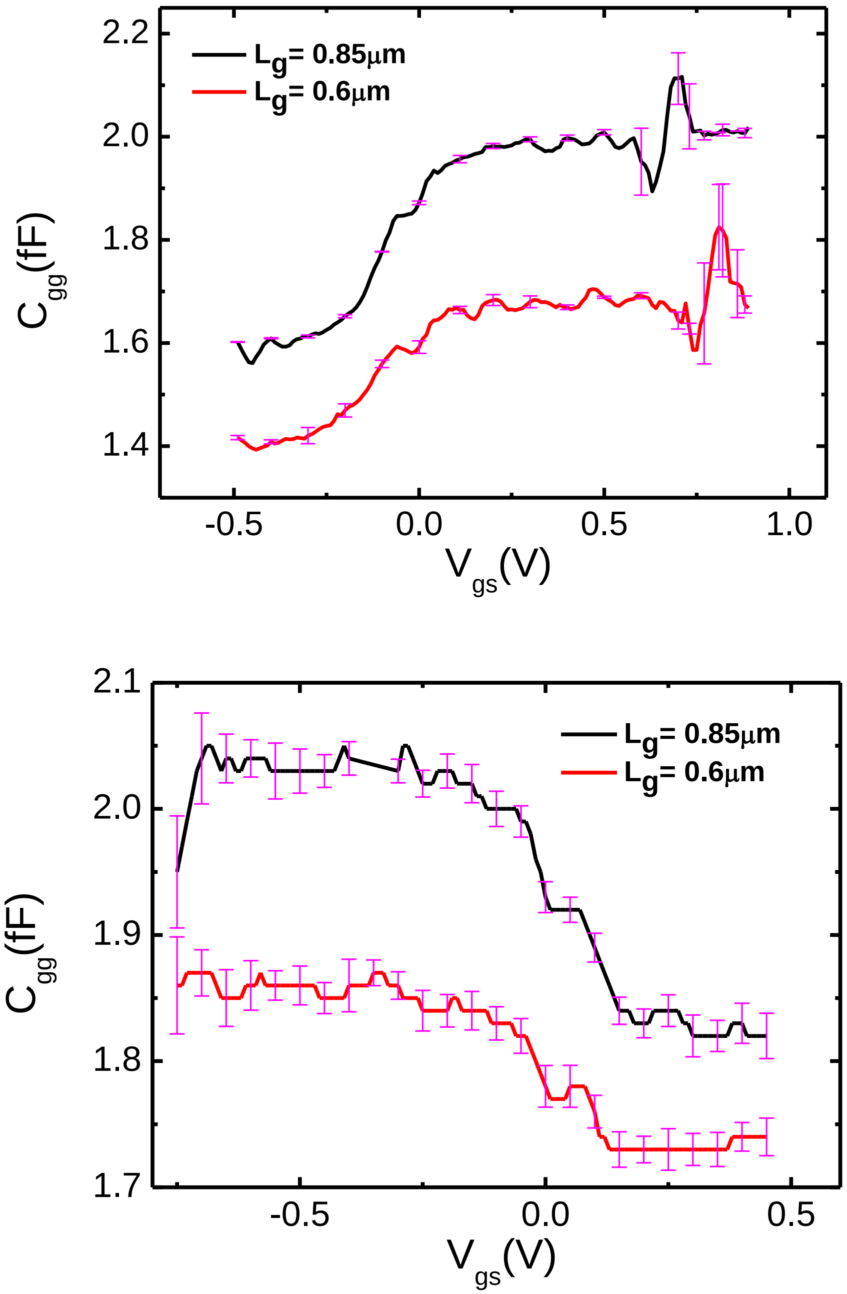


**Fig. 5-10:** Measured C-V characteristics of (a) N type and (b) P type DUT with gate lengths of 0.85 μm and 0.6 μm respectively. 16 measurements were taken on the same device and the error bars represent the standard deviation of capacitance data obtained from individual charge measurements.

**Fig. 5-10** shows the typical C-V characteristics for both n- and p-channel DUTs with two different effective gate lengths of 0.85μm and 0.6μm and a gate dielectric thickness of 6nm. The error bars represent the standard deviation of capacitance from 16-measurements after removing the outliers. The measured $C_{DUT}$ includes both the channel capacitance and the parasitic capacitance due to the DUT itself. Therefore, the minimum value of the $C_{DUT}$ measured is not zero. The minimum capacitance has two main components, namely, (i) the overlap and (ii) the fringing and overlap capacitance from gate to source/drain (S/D). We combined 2D and 3D simulation to estimate the contribution of each component (details elaborated in Chapter 6). The overlap capacitance and the fringing capacitance from gate to extension region of SNW combined is ~300 aF by 3D TAURUS simulation and the fringing capacitance from gate to S/D contact regions and contacts is ~100 aF by 2D simulation. Some polysilicon trapped underneath the SNW in the S/D extension region due to anisotropic etch contributes the rest.

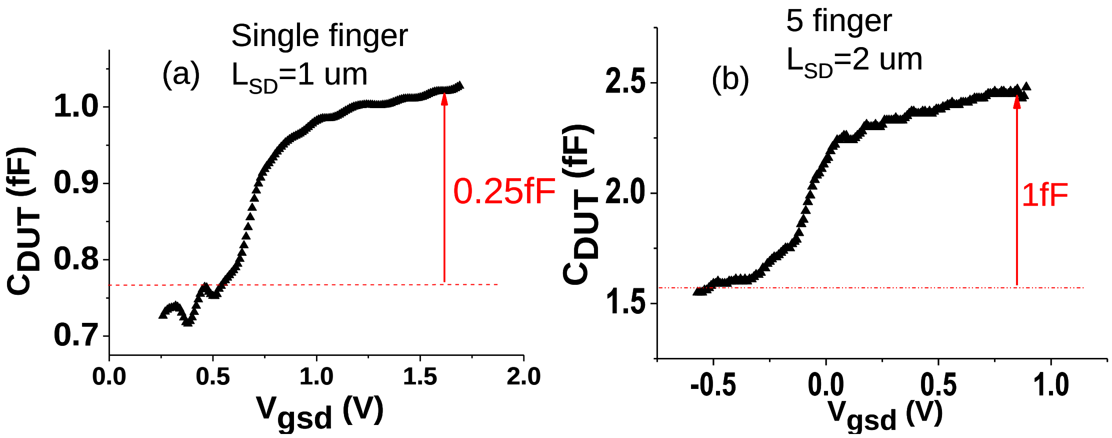


**Fig. 5-11**: $C_{DUT}$ –V curve for (a) single finger DUT with S/D contact region length of 1 µm and (b) 5 finger DUT with S/D contact region length of 2 µm, both extracted using CBCM.

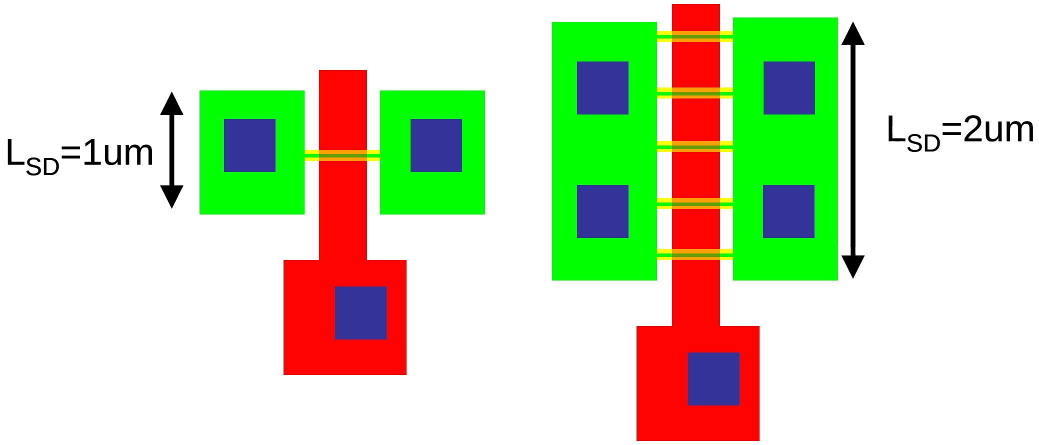


**Fig. 5-12**: Device layout schematic for single finger and 5 finger DUTs

This analysis of the capacitance can be illustrated further by comparing single finger and multiple finger NW $C_{DUT}$-V curves as shown in **Fig. 5-11**.The minimum capacitance, which is mainly due to overlap and fringing from gate to S/D contact regions, is 0.76 fF for single finger DUT and 1.53 fF for 5 finger DUT. It is directly proportional to the length of the S/D contact regions, which is 1 µm for single finger DUT and 2 µm for 5 finger DUT as shown by the layout schematics in **Fig. 5-12.** On the other hand, the channel capacitance, which only depends on channel length and cross-sectional area only, should be directly proportional to the number of fingers in the device. As shown **Fig. 5-11,** the channel capacitance, taken as the difference of the maximum and minimum value of $C_{DUT}$, is 0.25 fF for single finger DUT and 1 fF for 5 finger DUT. We believe this small discrepancy in terms of proportionality could be due to variation in wire diameters in single and multiple finger devices.

## *5.6 Chapter Summary and Conclusion*

The CBCM method has been successfully extended to measure sub-fF level bias-dependent capacitance of a single SNW transistor. Test keys with various numbers of fingers and gate lengths were designed and fabricated using the “Stringer-free” process described in Chapter 4. A test setup has been devised so that the C-V and I-V characteristics can be measured on the same NW DUT device. The CBCM scheme was benchmarked against C-V measurement results using conventional method on a DUT of higher capacitance before applied to single channel NW DUT.

**Table 5-2** compares the result of this work with the state-of-the-art technology in measuring NW capacitances. This is the first measurement of the femto farad scale single channel NW capacitance to our knowledge. The technique and the results are important to characterizing the intrinsic electrical properties of NW device and other three-dimensional nanoscale devices in general with little ambiguity.

| **Samsung<br>IEDM’07 [7]** | **MIRAI<br>IEDM’07 [10]** | **Uni. Of Tokyo<br>VLSI’08 [11]** | **Stanford (Tu)<br>Nano Lett.’07 [3]** | **This work** |
|---|---|---|---|---|
| 100x100 Array | ~500 Wires | 1000 Wires | Low-T, GeNW | 1 Wire,<br>room temperature |
| ~4-10 pF | N/A | 1-2 pF | ~1-2 fF | <1 fF |

**Table 5-2:** Comparing experiment results of CBCM technique with the state-of-the-art

## *Reference for Chapter 5*

[1] Y. Cui, Z. Zhong, D. Wang, W. U. Wang, and C. M. Lieber, “High performance Silicon Nanowire Field Effect Transistors”, *Nano Lett.*, vol. 3, no. 2, pp.149-152, 2003.

[2] T. Tesuka et al., “Observation of Mobility Enhancement in Strained Si and SiGe Tri-Gate MOSFETs with Multi-Nanowire Channels Trimmed by Hydrogen Thermal Etching”, *IEDM 2007*, pp887-890.

[3] R. Tu, Li Zhang, Y. Nishi, and H. Dai, “Measuring the Capacitance of Individual Semiconductor Nanowires for Carrier Mobility Assessment”, *Nano Lett.*, 7-6, pp1561-1565, 2007.

[4] J. C. Chen, B.W. McGaughy, D. Sylvester, and C. Hu, “An on-chip, attofarad interconnect charge-based capacitance measurement (CBCM) technique”, *IEDM Tech. Dig. 1996*, pp.69-72B.

[5] Y. W. Chang et al., “Charge-based capacitance measurement for bias-dependent capacitance”, *IEEE Ele. Dev. Lett. 27(3)*, pp.390-392, 2005.

[6] Y. W. Chang, H.-W. Chang, C.-H. Hsieh, H.-C. Lai, T.-C. Lu, W. Ting, J. Ku, and C.-Y. Lu, “A novel simple CBCM method free from charge injection-induced errors”, IEEE Elec. Dev. Lett. 25(5), pp.262-264, 2004.

[7] S. D. Kim, M. Li, Y. Y. Yeoh, K. H. Yeo, K. H. Cho, I. K. Ku, H. Cho, W. J. Jang, D. W. Kim, D. Park, and W. S. Lee, “Investigation of nanowire size dependency on TSNWFET”, *IEDM Tech. Dig. 2007*, pp. 891-894.

[8] B. Peirce, “Criterion for the Rejection of Doubtful Observations”, Astronomical Journal, II-45, pp.161-163, 1852.

[9] A. Savitzky & M. Golay, “Smoothing and differentiation of data by simplified least squares procedures” Analytical Chem., 36-8, pp. 1627-1639, 1964.

[10] T. Tezuka, E. Toyada, S. Nakaharai, T. Irisawa, N. Hirashita, Y. Moriyama, N. Sugiyama, N. Taoka, Y. Yamashita, O. Kiso, M. Harada, T. Yamamoto, and S. Takagi, “Observation of mobility enhancement in strained Si and SiGe Tri-Gate MOSFETs with Multi-Nanowire channels trimmed by hydrogen thermal etching”, *Tech. Dig. IEDM 2007,* pp. 371-374.

[11] J. Chen, T. Saraya, K. Miyaji, K. Shimizu, and T. Hiramoto, “Experimental study of mobility in [110]- and [100]-directed multiple silicon nanowire GAA MOSFETs on (100) SOI”, *VLSI Tech. Dig. 2008,* pp. 32-33, 2008.

# Chapter 6

# Device Analysis and Modeling for Single Channel NW Devices

## 6.1 Introduction

As transistor size shrinks down to the nanoscale as in the case of a nanowire (NW), the number of atoms in the cross section is drastically reduced to even countable numbers. In this regime, the atomistic property of the material can no longer be neglected: effects of atomistic strain, the underlying crystal symmetry, bond orientation and quantum mechanical confinement will affect the device operation and performance in a far more significant way [1]-[3]. Modeling and fitting of measurement data is the first and essential step towards the interpretation and understanding of experiment data as well as reliable device design at the nanometer scale. In the remaining part of this section, the simulation tools and techniques used in this chapter will be briefly introduced.

*COMSOL Multiphysics:* COMSOL Multiphysics is excellent, state-of-the-art software for the solution of many types of partial differential equations (PDE) by numerical techniques based on the finite element method of the spatial discretization. Its simulation environment facilitates all steps in three-dimensional (3D) modeling process – defining geometry, specifying physics, meshing and solving of PDEs. The ability to define and couple any number of arbitrary, nonlinear PDEs makes COMSOL Multiphysics a unique tool for sophisticated modeling applications [4]. COMSOL is used for simulation of 3D extrinsic capacitance of the NW devices in this study.

*Classical device simulation model:* Medici [5] models the classic two-dimensional (2D) distribution of potential and carrier concentration in a semiconductor device. It helps investigate internal device operations through potential, electric field, carrier and current density distributions by solving Poisson's equation and both the electron and hole current continuity equations. While the method deployed in Medici is classical, there is an additional quantum mechanical model in Medici which is capable of calculating the confined carrier distributions that occur near Si/SiO$_2$ interfaces – Modified Local Density Approximation (MLDA) [6]. It is based on a extension of the local density approximation to provide much improved computation efficiency. The MLDA model expressed the electron quantum potential as:

$$\psi_{qn} = \begin{cases} -\psi + \phi_n + V_T[\ln\left(N_c / n_{ie}\right) + F_{1/2}^{-1}\left(n_{MLDA} / n_c\right) & \text{Fermi} \\ -\psi + \phi_n + V_T \ln\left(N_c / n_{ie}\right) & \text{Boltzmann} \end{cases} \qquad (6\text{-}1)$$

where $\psi$ is the classical intrinsic potential, $\phi_n$ is the electron quasi-fermi potential and $n_{ie}$ is the effective intrinsic carrier density, $V_T$ is the thermal voltage and $F_{1/2}^{-1}$ is the inverse of the Fermi-Dirac integral, and $N_c$ is the conduction band density of states. The quantum potentials describe the deviation of the carrier concentrations from a classical distribution due to quantum mechanical confinement. The quantum potential is added to the classical intrinsic potential when calculating the carrier concentrations and solving Poisson and the continuity equations. Medici is used for modeling of channel charge, potential and intrinsic capacitance in this study.

*The sp$^3$ s*d$^5$ tight-binding model:* Due to the 2D quantum confinement, the bulk crystal symmetry is not preserved in NW structure any more. For this reason, atomistic bandstructure effects are expected to be important. While a full quantum mechanical transport model such as the non equilibrium Green's function (NEGF) [7] approach is very useful, it demands huge computational power and therefore is not feasible for a device with relatively large number of atoms. We adopted sp$^3$d$^5$s* tight-binding approach which offers to accurately model bandgaps (to within a few millielectron volts) and effective masses (to within a few percent) with an empirical bandstructure model. First, the energy dispersion (*E*-*k*) relations and their wavefunction is calculated by sp$^3$d$^5$s* tight-binding method [8]-[10]. Based on the computed *E*-*k* relations, the 2D Poisson equation can be solved in the cross section of the NW to obtain the electrostatic potential and special distribution of the charges.

Since the NW device we characterized has relatively long gate lengths (0.75 – 1 µm), the short channel effects are not playing any significant role in device performance. Therefore, a 2-D simulation model with sp$^3$d$^5$s* tight-binding method which capture the quantum effect within the nanowire cross section was adopted to reduce the requirement for the computation recourses in 3D simulation. On the other hand, though the quantum effects are of much importance for the calculation and modeling of the capacitance of nanowire, it is not essentially required for other parts of the device such as the parasitic capacitance calculation. To model the parasitic capacitances such as fringing and overlap capacitances with accuracy, the 3D structure of the NW device and the 3D distribution of the electric field need to be accurately captured. Therefore, a separate 3D device model was built in COMSOL [4] to simulate parasitic capacitances.

In this chapter, we report analysis and modeling of single channel NW devices based on electrical characterization using CBCM technique. First, a realistic three dimensional (3D) TCAD model with dimensions acquired from SEM and TEM images of the structure of the characterized device is built using COMSOL Multiphysics [4] to obtain the parasitic capacitance. Modeling of channel capacitance was first done in simple two dimensional (2D) model, and then in more sophisticated self-consistent sp$^3$s*d$^5$ tight-binding model [8]-[10]. This is perhaps the first report to present a comparison of the measured C-V data with carefully constructed simulation model of a single channel Si NW transistor. Finally, the series resistance and mobility were extracted based on the experimental results. These results are of vital importance for characterizing the transport and variability in the emerging research devices.

## 6.2 3D COMSOL Multiphysics Calculation of Parasitic Capacitance

As we have shown in **Fig. 5-9**, there is a capacitance floor in the C-V characteristic of the NW devices which is mainly due to the fringing and overlap capacitance. COMSOL Multiphysics, a state-of-the-art software package for 3D modeling was used to simulate and verify the experimental result.

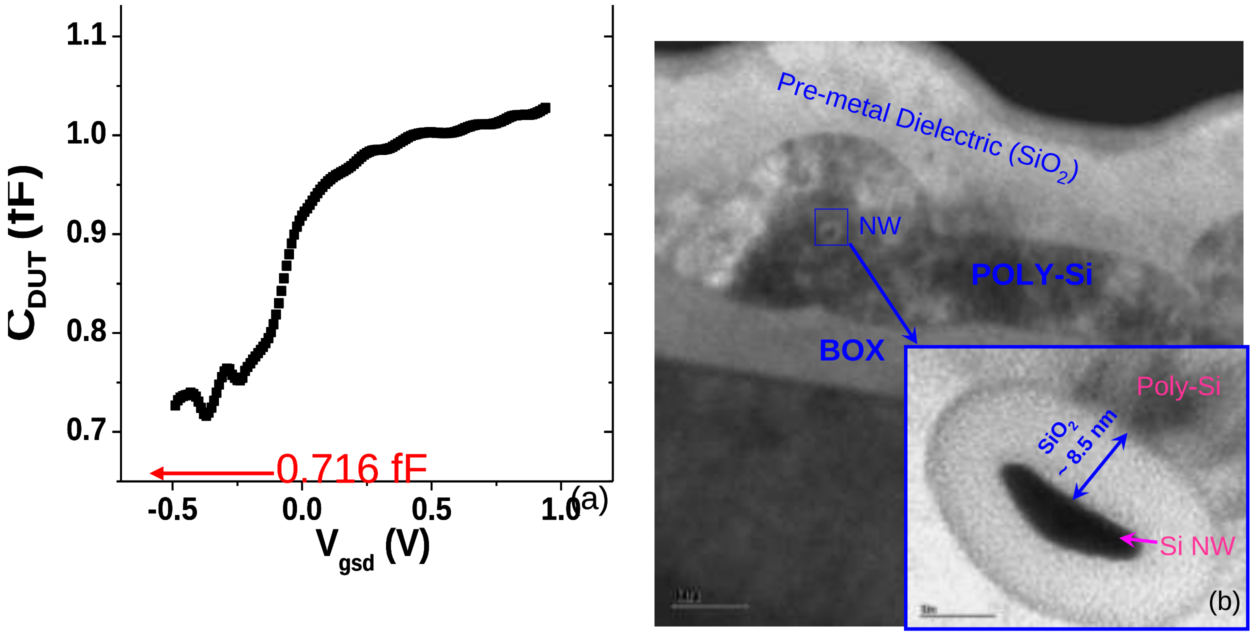


**Fig. 6-1:** (a) – the C-V characteristic of the NW device modeled (b) –cross sectional TEM view perpendicularly across channel of the medeled device, inset shows the detail for the NW channel

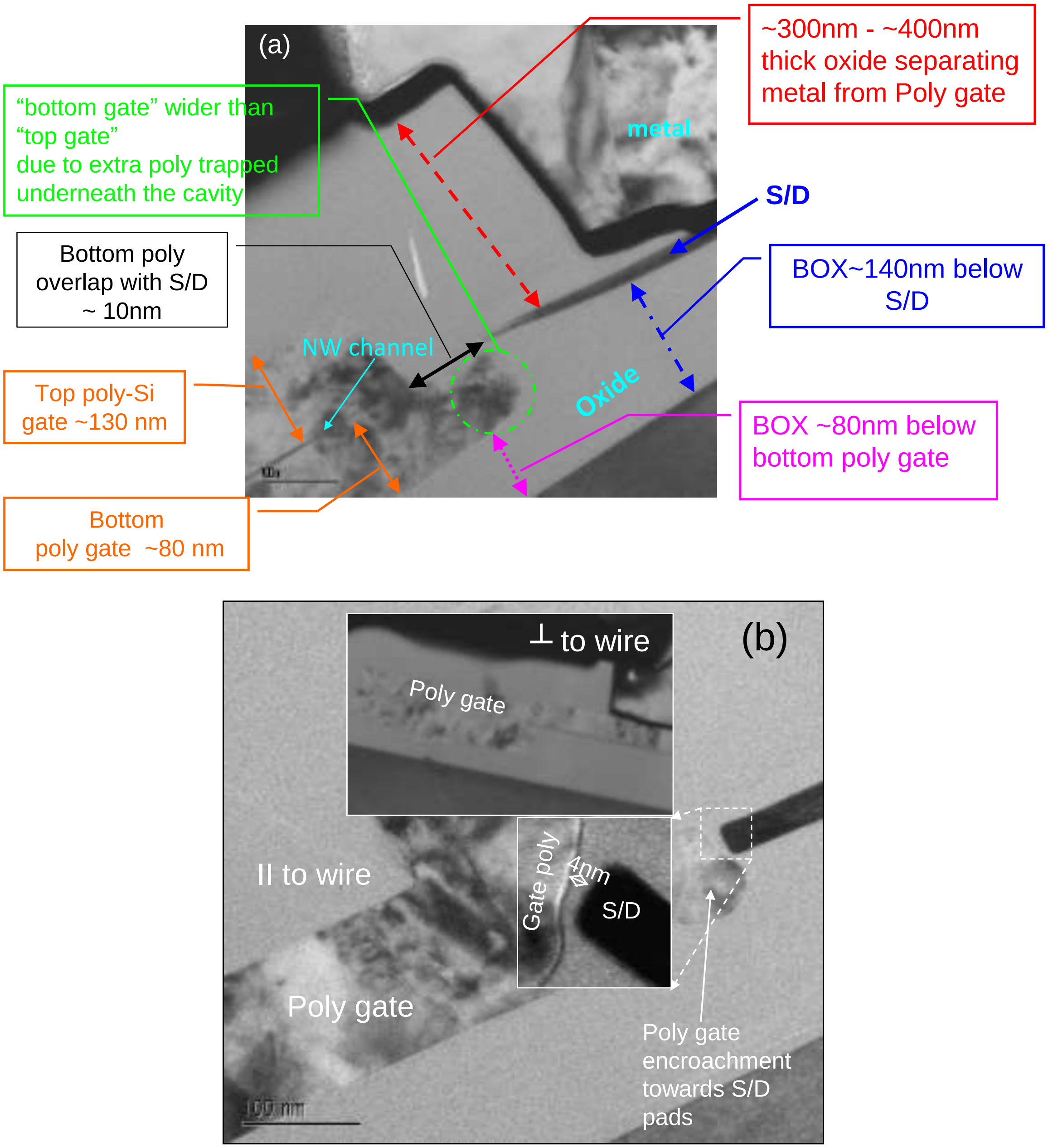

(a)
~300nm - ~400nm thick oxide separating metal from Poly gate
"bottom gate" wider than "top gate" due to extra poly trapped underneath the cavity
metal
S/D
Bottom poly overlap with S/D ~ 10nm
BOX~140nm below S/D
NW channel
Oxide
Top poly-Si gate ~130 nm
BOX ~80nm below bottom poly gate
Bottom poly gate ~80 nm
(b)
⊥ to wire
Poly gate
II to wire
Gate poly
4nm
S/D
Poly gate
Poly gate encroachment towards S/D pads
100 nm

**Fig. 6-2**: (a) – cross section TEM view along the channel direction of the modeled device, with dimensions of poly-Si gate, BOX layer, and pre-metal dielectric thickness illustrate (b) - The TEM cross-section in the direction of current flow shows that there is significant under-cut of the buried oxide during the local release of nanowires. This brings the gate poly to be as close to S/D contact area as the gate dielectric thickness increasing the parasitic $C_{G/S}$ as shown in the zoomed image in the inset.

The modeled NW device is an N-type NW device. TEM cross section images of the device were taken after electrical measurement. As shown in **Fig. 6-1** and **6-2**, the device has a triangular shaped channel cross section, gate dielectric thickness of around 8.5 nm and the minimum capacitance is 0.716 fF. The gate length of this device is 0.85 µm. It is important to note that in **Fig. 6-2** (a), the poly-Si were divided into half by the nanowire, and the bottom gate is wider than that of the top gate due to extra poly trapped underneath the cavity. However, the top and bottom poly layers are electrically connected. To calculate the capacitance correctly, these non-idealities in shape will need to be taken care of in the device structure. This can be seen more clearly in **Fig. 6-2** (b) which shows the cross section of poly-Si gate and the S/D of NW device in directions both along and perpendicular to the nanowire. Its inset shows that minimum oxide thickness between the S/D region and the bottom poly-Si region is as thin as 4 nm.

To investigate this further at reasonable level of complexity, we simulated the structure in two parts. For the first part, precise 3D device structure has been built in COMSOL to simulate the gate parasitic capacitance. Then the intrinsic gate capacitance is calculated using self-consistent $sp^3s^*d^5$ tight-binding model for density-of-state. Part of these works has been done by our collaborators in Purdue University. The subsequent part of this section addresses the first part of the capacitance investigation and the next section will address the second.

To make sure the accuracy of simulation, we first tested COMSOL with overlap and fringing capacitance of a planar device. A simple 2D MOSFET device structure with gate and source was built and the total extrinsic capacitance was benchmarked with the results reported in [11]. The device simulated is a P-type MOSFET device with gate height 0.4 µm, gate dielectric thickness 35 nm, junction depth 0.4 µm and overlap length 0.6 µm, as shown in **Fig. 6-3** (a). The total capacitance from gate to source simulated using the analytical model in COMSOL is matching well with the numerical result (**Fig. 6-3** (b)).

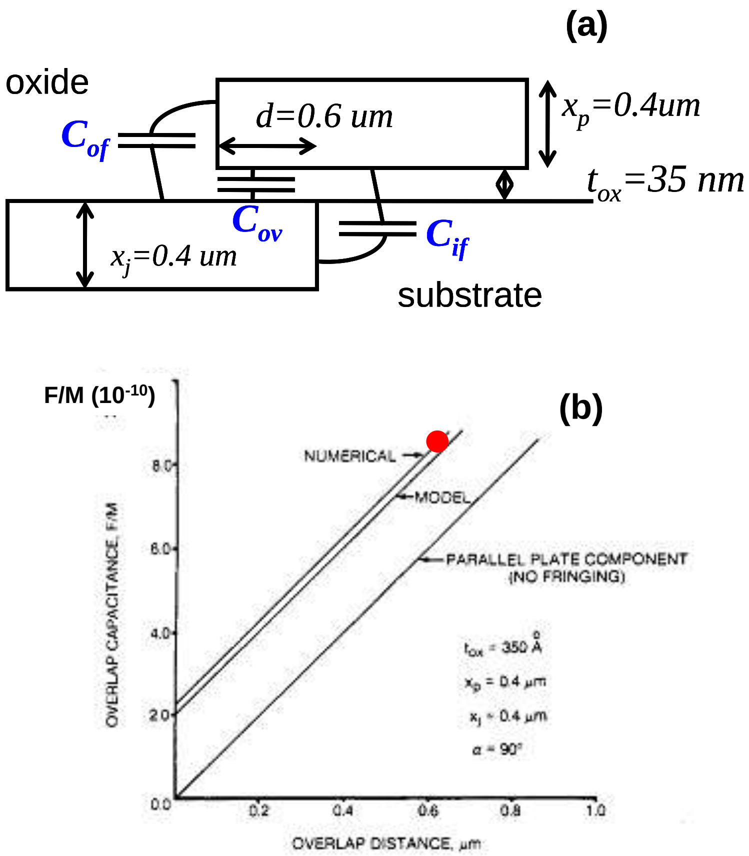


**Fig. 6-3**: (a) – the simulated device structure in COMSOL with dimension illustrated (b) – bench-marking the COMSOL analytical simulation result (marked by the red dot) with the results reported in [11].

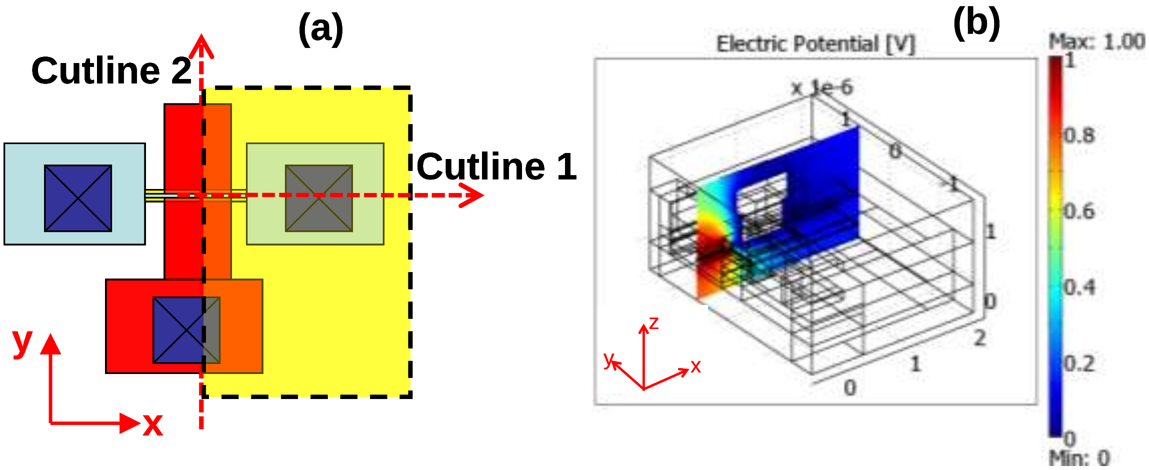


**Fig. 6-4**: (a) – a schematic of the device top view for 3D modeling of parasitic capacitance (not to scale). Only half of the structure is simulated due to symmetry (simulation domain is marked by yellow shading). (b) – A 3D snapshot of the simulation domain (ruler unit: μm). The cut plane parallel to z-direction shows the electric potential distribution in the structure.

Based on design layout and the SEM measurements made during fabrication, we constructed the simulation domain as shown by the area shaded in yellow in the schematic in **Fig. 6-4** (a). A 3D snapshot of the simulation domain in COMSOL is shown in **Fig. 6-4** (b), showing the electric potential along a cut plane in z-direction. Due to symmetry, only half of the device structure is simulated. Detailed device structure construction are shown in **Fig. 6-5,** which shows the electric potential distribution as well as dimensions along the two cut lines marked in **Fig. 6-4** (a).

As shown in both potential contour plots in **Fig 6-5**, the potential changes sharply near the corners of the poly silicon gate facing the S/D region, as well as in the narrow space separating the poly-Si gate contact region and the edge of the S/D region. Since channel shape is not particularly critical to parasitic simulation, the NW channel was simulated by a cylinder with diameter of 20 nm and the gate dielectric thickness is taken as 10 nm. The parasitic capacitance component in $C_{gg}$, including fringing and overlap, calculated using COMSOL is ~0.714fF which is very close to the measured minimum capacitance for the single finger device of the same gate length (as seen in **Fig. 6-1**). This simulation result not only further validates the CBCM measurement result, but also provides a fast and reliable model for estimation of parasitic capacitance in 3-D transistor structures using device cross sections.

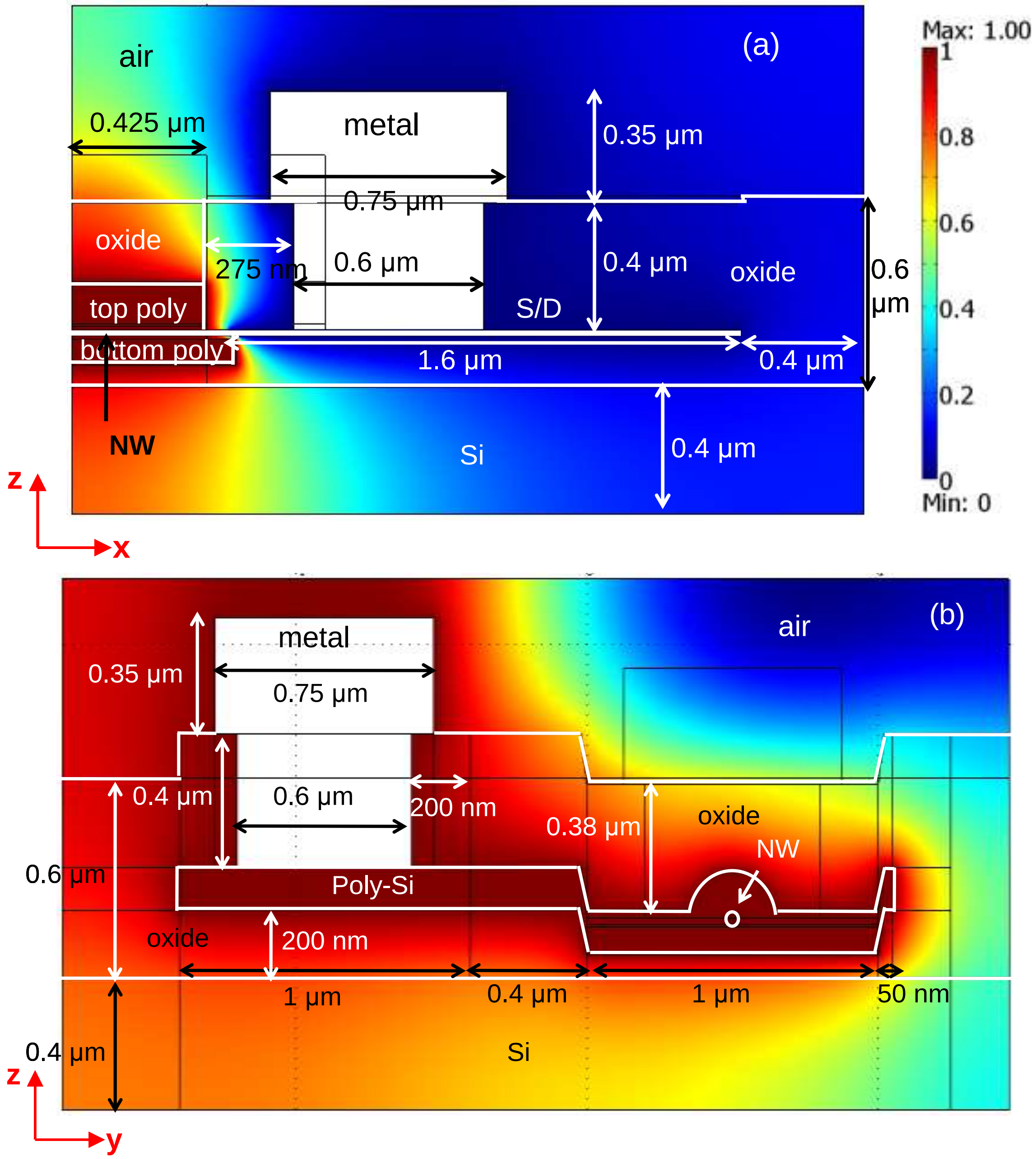


**Fig. 6-5**: Cross section view of the simulated device structure at bias $V_{gsd}$=1 V showing the electrical potential distribution and dimensions (a) – along *Cut line 1* (b) – along *Cut line 2* shown in **Fig. 6-3** (a). Both planes are along z-direction.

## 6.3 Simulation of Channel Charges and C-V

Two approaches were investigated for the simulation of the channel charges and intrinsic gate capacitance of the NW device. First one is using 2D simulation in Medici, which solves for charge and electron current using coupled Poisson and continuity equations.  In the second approach we use more sophisticated model: the Schrödinger and Poisson equations were solved self-consistently for the cross section of the NW; and the density of state was taken care of by $sp^3s^*d^5$ tight-binding model. Results of both will be discussed and compared in the remaining portion of this section.

### ***6.3.1 2D Medici model***

Medici offers a robust, fast and light weight option for simulating device characteristic in conventional device structures free of the influence of quantum mechanical effects. Therefore it is of particular interest to explore for NW deices how Medici results compares with other more sophisticated options.

In order to reflect the true electrostatic in the channel, a device structure was built to resemble the real cross section of the NW using coordinates taken from the actual device TEM cross section view.

**Fig. 6-6** (a) shows the cross section view of the simulation domain while (b) and (c) the electric potential as well as electron density simulated at bias $V_g$=1 V. Besides the gate electrode, another electrode was placed in the centre of the nanowire body. The total charge on the gate electrode was differentiated with respect to $V_g$ to obtain the gate capacitance. Because the channel is lightly doped, with the assumption that the majority carrier can be ignored, we only take the electron charge into account. This method leads to a low frequency C-V curve of intrinsic capacitance as shown in **Fig. 6-7**. It can be seen that after adjusting for the overlap and the bottom poly-gate region and parasitic capacitance (0.714 fF as seen in the previous chapter), the Medici simulated value is still less than measurement. The simulated intrinsic capacitance is less than the measurement value by 26.8%. Also note that there is a threshold voltage difference between the two curves.

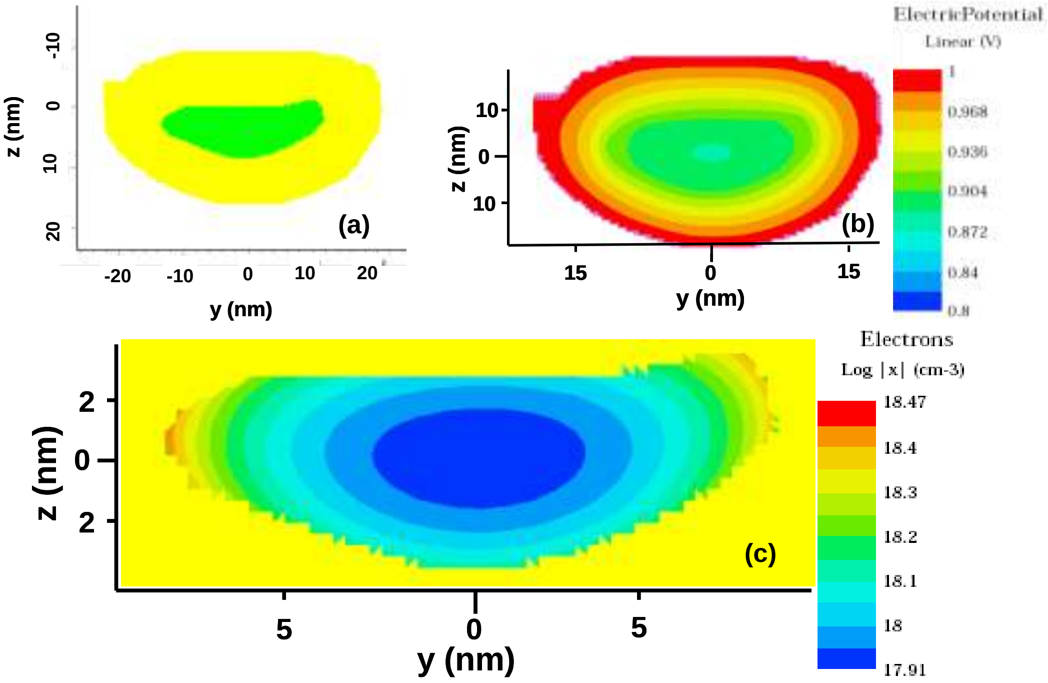


**Fig. 6-6:** (a) – Simulation domain built in Medici according to the TEM cross section image of the NW device (b) and (c) – Medici simulation results of electron potential (in Volt) and electron charge density (in cm$^{-3}$) respectively. Bias condition for (b) and (c) is $V_g$=1 V.

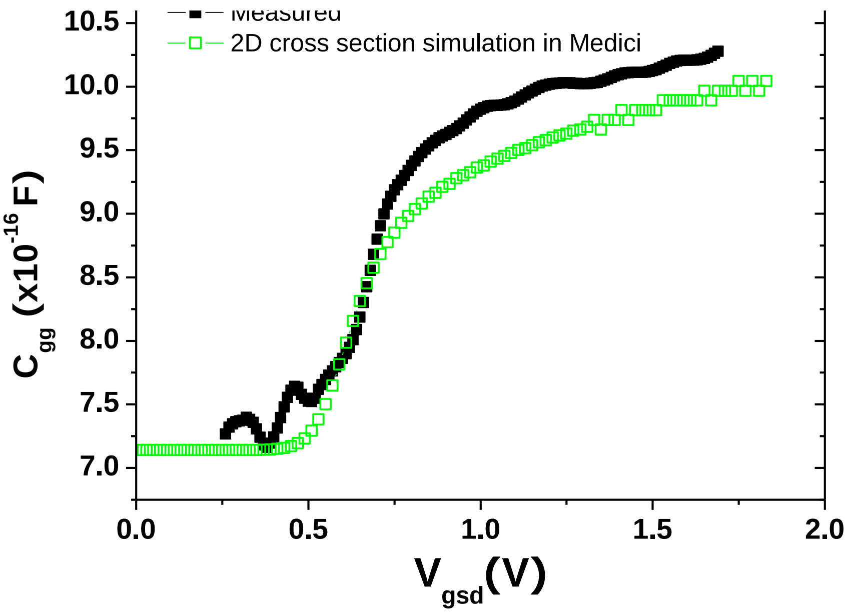


**Fig. 6-7**: Comparing the measurement C-V curve and the C-V extracted from Medici simulations.

### 6.3.2 *Self-consistent simulation using sp$^3$s*d$^5$ tight-binding model*

For extremely scaled device dimensions, the crystal symmetry, bond orientation and quantum mechanical confinement will matter in transport and device operation. These are the atomistic effects which the usual effective mass approximation (EMA) fails to capture, atomistic modeling is therefore a favored option in understanding the electrical characteristic of the ultra-small cross section NW devices.

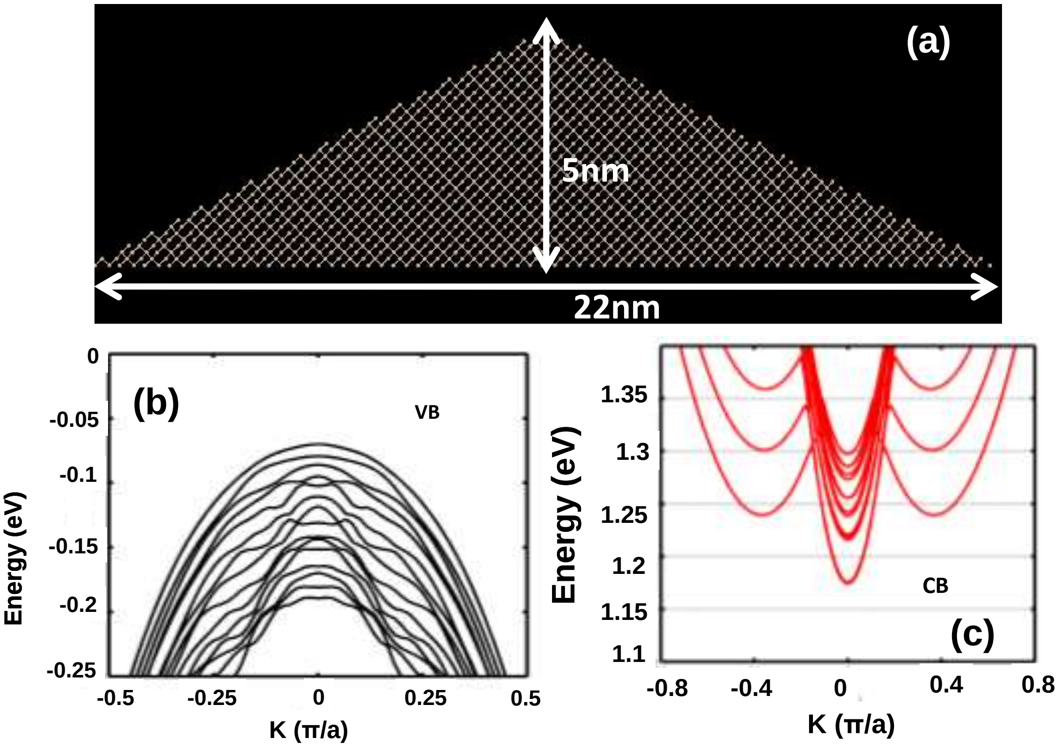


**Fig. 6-8**: (a) – The lattice in the wire cross section in (100)/[100] direction (b)&(c) – showing the valence and conduction bandstructure respectively for the NW under consideration

The simulation procedure is as follows: First the bandstructure of the wire is calculated using an atomistic tight-binding model. In this case, each atomic side in the zincblende lattice is represented by a sp$^3$d$^5$s* basis in the Hamiltonian, as shown in **Fig. 6-8**. The atoms at the surface are passivated in the sp$^3$ hybridization scheme [8].

Schrödinger equation and Poisson equation are solved self-consistently over the cross section to obtain the electrostatic potential at different bias conditions, and in turn, the intrinsic capacitance per unit length of the wire can be obtained. Note that in this case, the two contacts were placed on the outer surface of the $SiO_2$ dielectric layer and the center of the Si NW channel.

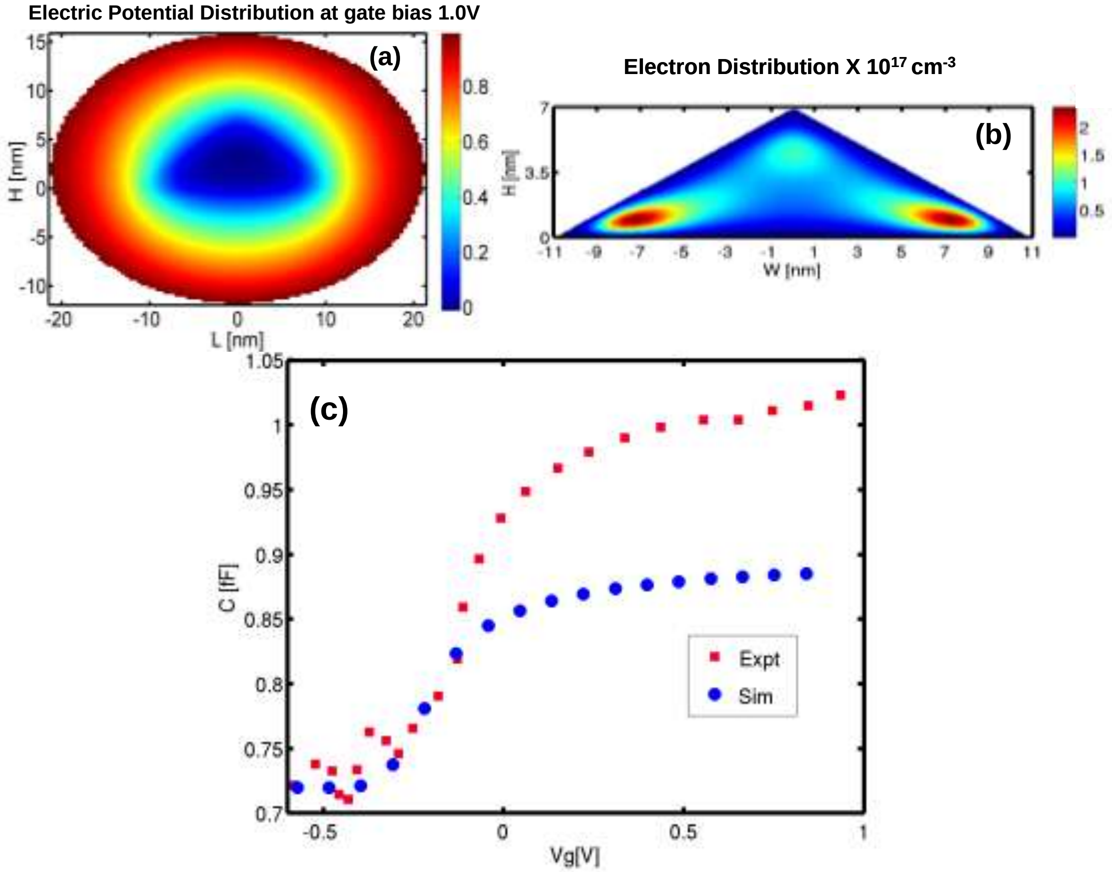


**Fig. 6-9**: The simulated (a) – electric potential distribution and (b) – electron charge distribution in a triangular NW cross section (c) – the C-V curve obtained by measurement and the simulated C-V curve (the parasitic capacitance simulated by COMSOL was added to the intrinsic gate capacitance)

A simple triangular model assimilating the perimeter and area of the TEM cross section image of the measured device, as shown in **Fig. 6-1** (a) was built. The cross section Si surface orientation is (100)/[100] and it has a cross section area of 50 nm$^2$. Shown in **Fig. 6-9** are the simulated potential, electron distribution plot and the C-V curve obtained using self-consistent sp$^3$s*d$^5$ calculation. Comparing **Fig. 6-9** (a) and **Fig. 6-6** (b), it can be seen that the body potential of the NW channel was much overestimated in the Medici simulation, probably due to the lack of a body contact. **Fig. 6-8** (b) shows higher electron concentration near the corner of the triangle which is not captured by Medici. However, the simulated intrinsic capacitance is smaller than the measurement value by as much as 0.15 fF or 40% of the measurement.

A second 2-D model for atomistic description of the device cross section is built according to the shape and dimensions of Si and dielectric obtained from TEM cross section image of the measured device, as shown in **Fig. 6-1** (b). This is very different from the most of usual approach which simplifies the cross section with a triangular, circular [11] or a rectangular [10] shape. We find this very important since the electrostatic potential variation is a critical step in evaluating the NW capacitance. Although this factor may not cause huge inaccuracy in currents, it strongly influences the magnitude and placement of charge in the cross section of the channel. Therefore, the for valid estimation of the channel charge, and thus intrinsic capacitance, the precise cross section shape and size must first be captured.

The simulated electrostatic potential and the electron charge distribution with the NW channel cross section of precise shape are shown in **Fig. 6-10** (a) and (b). Comparing them with Fig. 6-8, we can see that not only the potential distribution contour is clearly affected by the shape of channel, but the peak electron charge density is much higher in **Fig. 6-10** (b). It can be clearly observed that the positions of the peak of electron density are shifted much towards the Si-SiO$_2$ interface in **Fig. 6-10** (b) as compare to that in **Fig. 6-9**. Also, the channel charge density overall is much higher than in **Fig. 6-9** (b), showing that the electrical property of NW is not only sensitive to quantization size but also quantization shape.

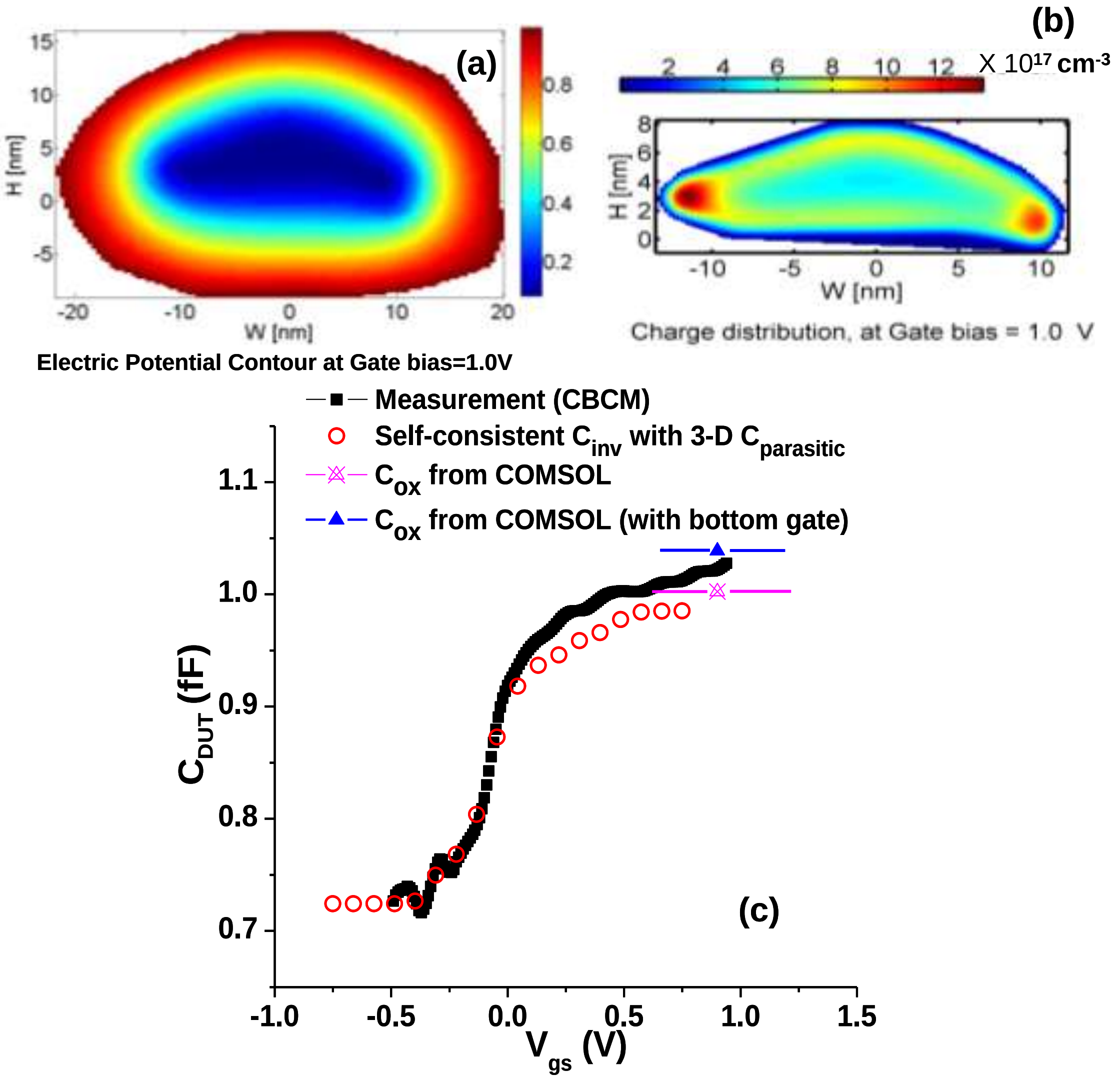


**Fig. 6-10** : (a) and (b) Potential and electron density in channel obtained by self-consistent computation, respectively (c) – measurement C-V curve compared against the simulated C-V with $sp^3s^*d^5$ tight-binding model with precise channel cross section dimension (the parasitic capacitance simulated by COMSOL was added to the intrinsic gate capacitance)

Shown in **Fig. 6-10** (c) is the C-$V_{gs}$ curves of single channel Si NW transistor measured by the CBCM technique and self-consistent intrinsic Si NW simulated with $sp^3s^*d^5$ tight-binding model added with the 3D electrostatic parasitic capacitance simulated in COMSOL. It shows very good agreement of the measured $C_{DUT}$ with the simulated values. The simulated intrinsic inversion gate capacitance is about 0.261 fF, very close to the measured intrinsic inversion capacitance ~0.296 fF (there is some ambiguity in this estimation due to the uncertainty of the minimum capacitance is the measured C-V curve).

## *6.4 Mobility Extraction and Analysis for Single Channel SiNW Device*

Accuracy in the simulated/measured capacitance is extremely important for transport characterization because it is directly related to the charge in the channel at a given bias. The C-V curve obtained from single-channel NW device enables evaluation of the electron and hole mobility in NW using experimentally determined charge data. The results are important to the understanding of the intrinsic electric properties of Si NW with little ambiguity and it can also shed light into optimization of Si NW devices.

To obtain the channel resistance needed of mobility extraction, the total resistance was first extracted from $I_d$-$V_g$ curve as according to **Eqn. 6-2** and shown in **Fig. 6-11.**

$$R_{tot} = \frac{V_{ds}}{I_{ds}} \qquad \textit{(6-1)}$$

The channel resistance is than corrected for the series resistance according to **Eqn. 6-2** and **6-3** by plotting the total resistance ($V_{ds}/I_{ds}$) vs $1/(V_{gs}$-$V_{th}$-$V_{ds}/2)$ at high gate voltages.  The series resistance can then be derived by taking the intersect point of the line on the y-axis as shown in **Fig. 6-12**. Here β is the gain parameter ($\mu C_{ox}W/L$), $R_S$ is the total series resistance and other symbols have their usual meaning.

$$I_{ds} = \mu C_{ox}\frac{W}{L}(V_{gs} - V_{th} - \frac{1}{2}V_{ds})(V_{ds} - I_{ds}R_s) \qquad \textit{(6-2)}$$

$$\frac{1}{\beta(V_{gs} - V_{th} - \frac{1}{2}V_{ds})} = \frac{V_{ds}}{I_{ds}} - R_s \qquad (6\text{-}3)$$

The series resistance is 150 kΩ and 224 kΩ for N and P-type NW device respectively. This significant difference in the series resistances for N and P-type DUT devices is attributed to difference in carrier mobility, possible device-to-device dimensional variation and doping level differences.

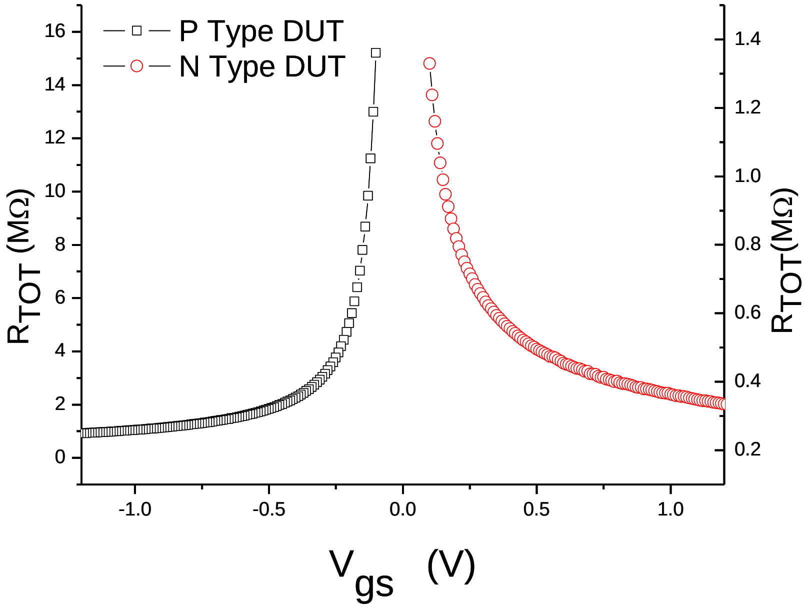


**Fig. 6-11**: Plotted total resistance ($R_{tot}=V_{ds}/I_{ds}$) showing the asymptotic behaviour of $R_{tot}$ vs $V_{gs}$ for both N and P-type DUT NW devices

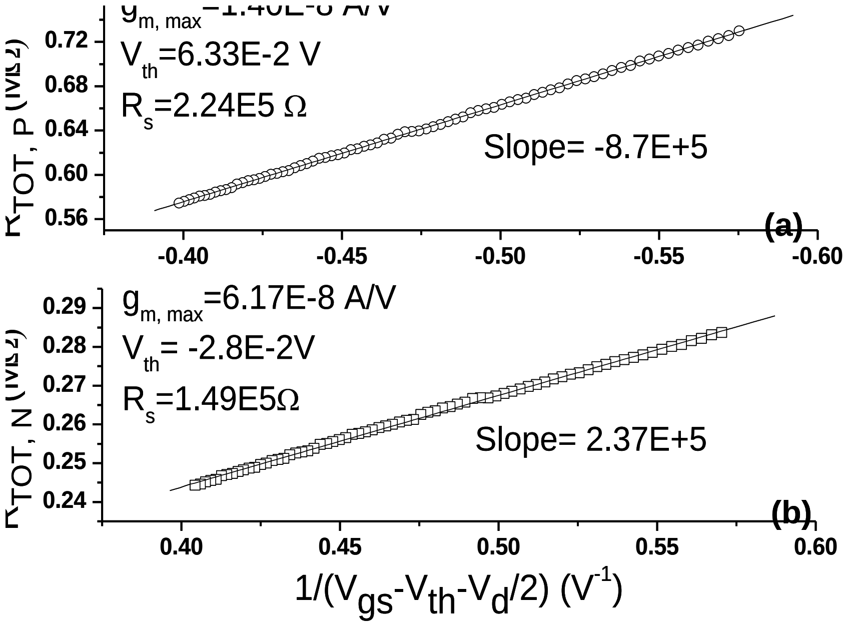


**Fig. 6-12**: Series resistance for (a) – P-type and (b) – N-type Si NW channels extracted from the linear region of Id-$V_{gs}$ characteristics at high $V_{gs}$ by plotting ($R_{tot}$=$V_d$/$I_d$) against 1/($V_{gsd}$-$V_{th}$-$V_d$/2).

If the transverse electric field dependence is taken into account, we have:

$$\frac{\mu}{\mu_0} = \frac{1}{\left(1+\left(\frac{E_t}{E_{CT}}\right)^{\gamma}\right)^{1/\gamma}} \qquad (6\text{-}4)$$

Where $E_t$ is the transverse electric field and $E_{CT}$ is the critical transverse electric field. With Eqn. (6-4), the series resistance equation can be modified as follows:

$$\frac{1+\theta(V_{GS}-V_{TH}-I_{DS}\frac{R_S}{2})}{(V_{GS}-V_{TH}-\frac{V_{DS}}{2})} = -\beta_0(\frac{V_{DS}}{I_{DS}}-R_S) \qquad (6\text{-}5)$$

For different values of θ, Eqn. (6-5) was linearly regressed iteratively and relationship between θ and $R_S$ is shown in **Fig. 6-13**. The convergence was obtained in just 2-3 iterations. This suggests that when we remove the series resistance based on a single device, the effect of normal field also gets de-embedded explaining the flatness of mobility values in strong inversion. We deliberately avoided combining the I-V data of two devices as any likely difference in the channel cross-section area may seriously affect the extracted mobility value.

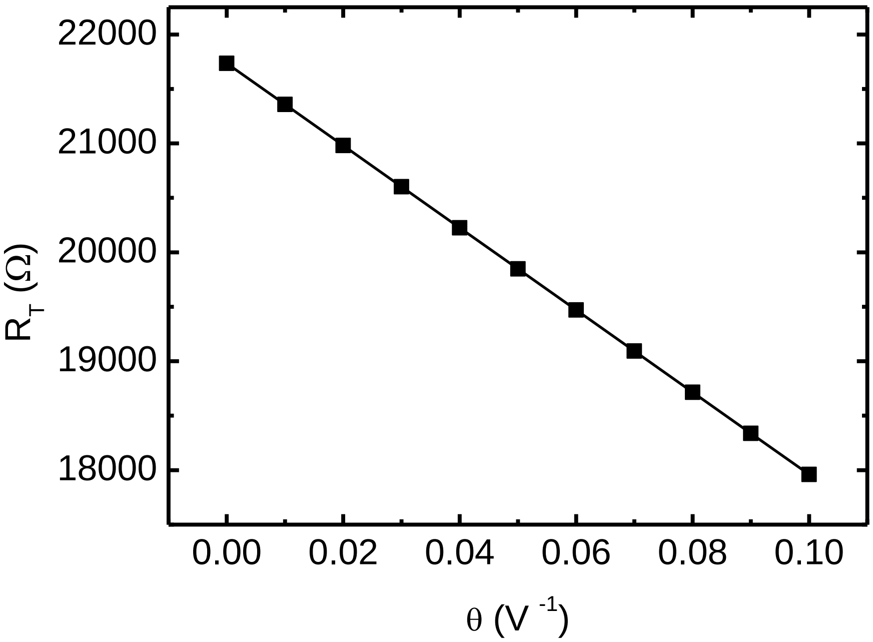


**Fig. 6-13**: Extracted series resistance for different value of θ from a single device showing that θ and series resistances trade-off with each other.

Carrier mobility can be extracted using the channel current in the linear region (|Vds|=10 – 50mV) to give the channel resistance and the channel carrier charge obtained by integrating the C-V curve for each gate voltage starting from the threshold voltage assuming voltage independent fringing and parasitic capacitances. The mobility is calculated using

$$\mu = \frac{L^2}{Q(V_{GS}) * R_{chnl}(V_{GS})} \qquad \textit{(6-4)}$$

where $L$ is the effective gate length (defined by reverse gate window), $R_{chnl}$ is the channel resistance and $Q$ is the total charge in the channel.

**Fig. 6-14** (a) and (b) show the effective mobility extracted for N-type NW device. The average mobility values of electrons is ~600 cm$^2$/V-s and ~250 cm$^2$/V-s for holes –

both higher than those reported for unstrained silicon nanowire channels. Chin et al. [13] report a mobility value of 450 cm$^2$/V-s and Hashemi et al. [14] report ~250 cm$^2$/V-s for unstrained GAA transistors. Interestingly, the mobility values are flat as a function of gate voltage. This is because we have extracted the series resistance value from the total measured resistance of the transistor in linear region explained by Eqn. 6-5 and Fig. 6-12 We have assumed a constant value of θ, which may not be true with varying bias conditions. Both channel length were assumed to be 0.85 µm. The sources of uncertainties in these mobility values include: the uncertainties in the characteristics of fringing and parasitic capacitance (e.g. voltage dependence) and change in gate length on account of process variations.

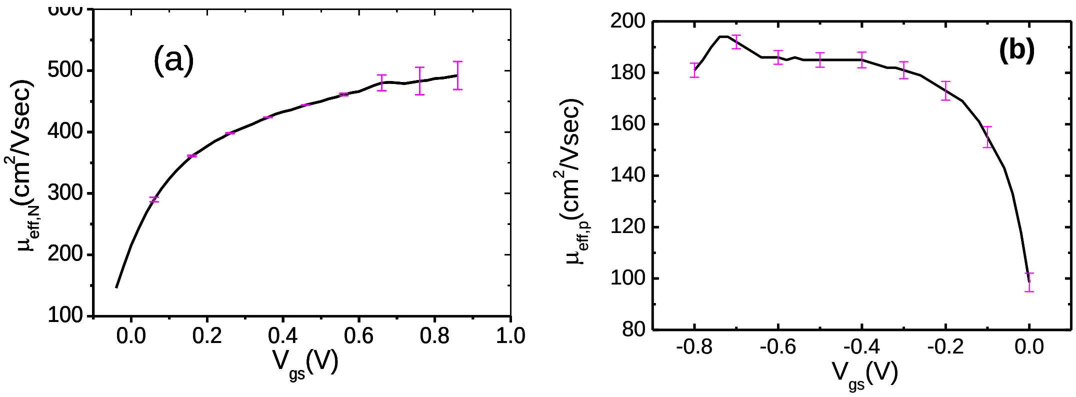


**Fig. 6-14**: (a) – Inversion charge density plotted against $V_{gs}$ (b) – effective mobility extracted vs. inversion charge density for P-type NW device.

**Fig. 6-15** shows the measured mobility distribution for two gate dielectric thickness for both N and P type devices. It be seen that the devices with 10 nm gate oxide has average mobility higher than that of 6 nm gate oxide by approximately 1.5 times. This could be on account of reduction in the roughness at the interface during longer oxidation times. The other plausible reason could be stress generated in the nanowire channel during longer gate oxidation. Interface roughness has been proposed to strongly influence the carrier mobility [13] in nanowires also.

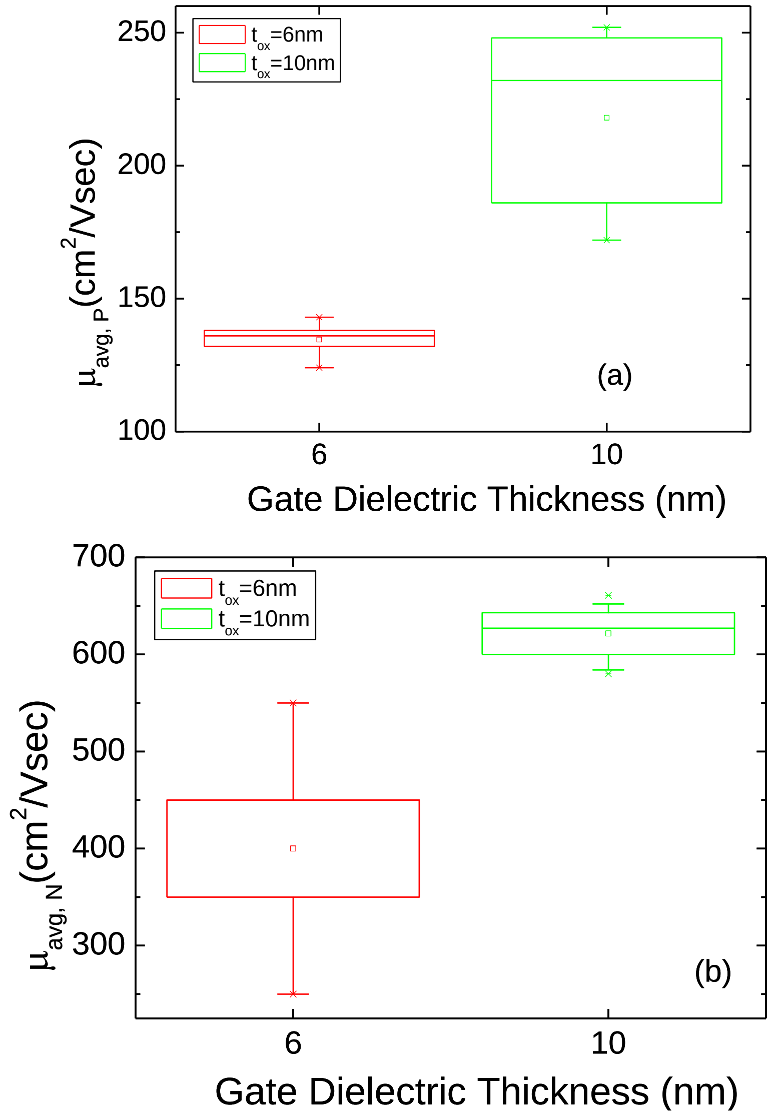


**Fig. 6-15**: Box plot with distribution of the average strong inversion mobility of the (a) – N and (b) – PMOS transistors respectively for two different thermally grown gate dielectric thicknesses. The mobility reduces for devices with lower gate dielectric thickness on account of more significant surface roughness scattering.

## *6.5 Chapter Summary and Conclusion*

In this chapter, we reported modeling of intrinsic and extrinsic device capacitances using combination of 2D and 3D simulations. The parasitic capacitance very close to the measured value has been calculated by building a 3D device model based on TEM images and device dimension measurements. The channel intrinsic capacitance was calculated using both Medici and the more sophisticated self-consistent $sp^3s^*d^5$ tight-binding bandstructure model. It was found that the 2D Medici simulation failed to simulate the channel potential accurately and therefore the inversion charge distribution also cannot be calculated in reliably. A good fit of the C-V curve can be obtained by elf-consistent $sp^3s*d^5$ tight-binding model but only with the exact cross section following the TEM image of the NW cross section, for it was found that the electrical property of the NW channel cross section is sensitive to quantization shape. Therefore, for modeling of nanoscale capacitance of the NW devices, the atomic simulation is necessary and care must be taken to the reproduce the channel shape as well for accurate result.

The electrical characterization obtained from experiment was used for effective mobility extraction for N and P-type NW channels. Major uncertainties and source of errors in capacitance measurement and mobility extraction are identified and discussed.

## *Reference for Chapter 6*

[1] H. Majima, Y. Saito, and T. Hiramoto, "Impact of quantum mechanical effects on design of nano-scale narrow channel n- and p-type MOSFETs", *IEDM Tech. Dig.*, pp. 951-954, 2001.

[2] J. Wang, A. Rahman, G. Klimeck, and M. Lundstrom, "Bandstructure and orientation effects in ballistic Si and Ge nanowire FETs", *IEDM Tech. Dig. 2005,* pp. 530-533, 2005.

[3] N. Singh, F. Y. Lim, W.W. Fang, S. C. Rustagi, L. K. Bera, A. Agarwal, C. H. Tung, K. M. Hoe, S. R. Omampuliyur, D. Tripathi, A. O. Adeyeye, G. Q. Lo, N. Balasubramania, and D. L. Kwong, "Ultra-narrow silicon nanowire gate-all-around CMOS devices: Impact of diameter, channel orientation and low temperature on device performance", *IEDM Tech. Dig.* , 2006, pp. 548-551.

[4] COMSOL Multiphysics [Online], available at http://www.comsol.com

[5] Taurus Medici TCAD Tools [Online], available at http://www.synopsys.com/Tools/TCAD/DeviceSimulation/Pages/TaurusMedici.aspx

[6] G. Paasch and H. Ubensee, "A modified local density approximation: electron density in inversion layers", *Phys. Stat. Sol. B*, vol. 113, pp. 165-178, 1982

[7] S. Datta, Non-Equilibrium Green's Function (NEGF) Formalism: An Elementary Introduction, *Proceedings of IEDM, 2002*, pp. 703-706

[8] G. Klimeck, F. Oyafuso, T. B. Boykin, R. C. Bowen, and P. von Allmen, "Development of a Nanoelectronic 3D (NEMO 3D) Simulator for Multimillion Atom Simulations and Its Application to

Alloyed Quantum Dots" (INVITED), Computer Modeling in Engineering and Science (CMES), vol 3, no. 5, pp 601-642, 2002

[9] Paul, Abhijeet; Luisier, Mathieu; Neophytou, Neophytos; Kim, Raseong; McLennan, Michael; Lundstrom, Mark; Klimeck, Gerhard (2006), "Band Structure Lab," doi: 10254/nanohub-r1308.6

[10] N. Neophytou, A. Paul, M. S. Lundstrom, G. Klimeck, "Bandstructure Effects in Silicon Nanowire Electron Transport," *IEEE Transactions on Electron Devices*, vol.55, no.6, pp.1286-1297, June 2008

[11] R. Shrivastava and K. Fitzpatrick, “A simple model for the overlap capacitance of a VLSI MOS device”, ," *IEEE Transactions on Electron Devices*, vol.29, no.12, pp.1872-1875, 1982.

[12] R. Tu, Li Zhang, Y. Nishi, and H. Dai, “Measuring the Capacitance of Individual Semiconductor Nanowires for Carrier Mobility Assessment”, *Nano Lett*., 7-6, pp1561-1565, 2007.

[13] J. Chen, T. Saraya, and T. Hiramoto, ”Electron mobility in multiple silicon nanowires GAA nMOSFETs on (110) and (100) SOI at room temperature and low temperature” , *IEDM Tech Dig.* 2008, pp. 1-4.

[14] P. Hashemi, L. Gomez, M. Canonico, and J. L. Hoyt, ”Electron transport in Gate-all-around uniaxial tensile strained Si nanowire n-MOSFETs ” *IEDM Tech dig.* , 2008, pp.1-4.

# Chapter 7

# Conclusions and Future Work

## 7.1 Summary and Major Contributions of Dissertation

FinFETs and nanowire (NW) FETs are emerging as leading contenders of next generation electron devices. These devices exhibit excellent control of SCE which is necessary for continuing scaling of the gate length. This dissertation focuses on silicon nanowire transistors fabricated using conventional CMOS platform. Its main purpose was to overcome major challenges in fabrication, capacitance measurement and characterization, as well as modeling.

Chapter 2 addressed the scaling issues related to ultra-thin body multi-gate devices by extensive simulations using three dimensional (3D) Taurus© process and device tools. FinFET devices with under-lapped gate structure were chosen as object for the study for its superior control of short channel effect and its technology compatibility. It is found that the 3-D device structural parameters such as *spacer width and material*, *gate electrode thickness* as well as *fin pitch* have significant effect for under lapped gate devices through 3D fringing electric field. Performance impact of geometry factors was carefully analyzed by comparing channel electrostatic as well as the fringing capacitance resulting from changes in geometry. Scaling down of the high-κ dielectric thickness is found favorable for device performance with particular reference to suppression of SCEs and switching speed even though the enhancement in the drive current is relatively moderate. Further, the adoption of relatively thinner spacers of higher permittivity material is found to be beneficial to device performance due to the suppression of SCE.

Very thin and highly conductive metal gate might be needed to reduce the gate-to-drain and gate-to-substrate fringing capacitances for faster switching. Also, it was found that the large value of the ratio of fin pitch/gate electrode thickness has an adverse effect on the performance of multi-fin devices. These findings and conclusions lead to important recommendations for the design and optimization of transistors of future generation. They are not applicable to FinFETs but also to other multi-gate ultra-thin body devices such as Omega FETs, nanowire FETs and nanotube transistors.

As the gate capacitance of devices decreases with scaling and the role of the parasitic capacitance becomes more significant, characterization of various capacitance components is becoming more important. As devices scale down into the nano-meter regime, the conventional measurement techniques and capacitance models are no longer sufficient due to their 3-D structure and the variability in nano-scale. In Chapter 3, we reviewed the charge based capacitance measurement (CBCM) as a simple and effective solution for femto-farad scale capacitance characterization. The three varieties of the CBCM setup schemes, reported in literature, were discussed in detail and were effectively evaluated through mixed device and circuit mode simulation in Medici. The impact of charge injection in CBCM technique were discussed and highlighted. For the practical application of CBCM to measurement, the main sources of errors in CBCM were identified and investigated, and solutions were suggested. In addition, test key design guideline which require that the parasitic capacitance to be less than twice of the estimated DUT capacitance was suggested for minimizing errors in CBCM.

In order to test the performance of CBCM measurements on real nano-scale devices, test keys were fabricated with nano-wire device as DUTs. The fabrication process for

NW FET device on a platform compatible with conventional CMOS technology were described in detail in Chapter 4. The critical steps were identified and solutions to some of very critical technological steps, such as stess-limited oxidation, lithography, and device structure free from poly-silicon stringer were suggested and implemented. This result implies that the NW FETs, with a process of enhanced controllability, is a step closer to efficient production with the state-of-the-art semiconductor technology.

In Chapter 5, we present the design of CBCM test key circuit and measurement scheme for obtaining I-V and C-V from the same single channel NW device. This is perhaps the first demonstration of successful measurement of femto-farad scale voltage dependent NW capacitance at room temperature. The CBCM method has also been validated by benchmarking its result with that of calibrated conventional CV meter for a NW device having larger capacitance showing good accuracy.

With accurately measured capacitances from single NW devices, the intrinsic and parasitic capacitances of nanowire devices can be analyzed and modeled in detail. In Chapter 6, we reported modeling of intrinsic and extrinsic device capacitances using combination of 2D and 3D simulations. The parasitic capacitance very close to the measured value has been calculated by building a 3D device model using COMSOL. The channel intrinsic capacitance was calculated using both Medici and the more sophisticated self-consistent $sp^3s^*d^5$ tight-binding model. It was found that the 2D Medici simulation failed to simulate the channel potential accurately and therefore the inversion charge distribution also cannot be calculated reliably. A good fit of the C-V curve can be obtained by self-consistent $sp^3s^*d^5$ tight-binding model but only with the exact cross section following the TEM image of the NW cross section, for it was found that the shape of the channel cross section affects the charge distribution significantly. This report is perhaps the first one to present a comparison of the measured C-V data with carefully constructed simulation model of a single channel NW transistor. The results are of vital importance for characterizing the transport and variability in the emerging research devices. Also, the electrical characterization obtained from experiment was used for effective mobility extraction for N and P-type NW channels. Record high carrier mobility for both electrons and holes were reported.

## *7.2 Suggestions for Future Works*

This work can be extended as follow in the future:

### *7.2.1 Self-limited oxidation modeling*

We have found clear evidence of self-limited oxidation of nanowires in the experiment results presented in Chapter 4. However, the phenomenon cannot be reproduced in process simulations with the existing models. For the same reason, the self-limited oxidation is only explained qualitatively in this work. Since self-limiting oxidation can be an effective method to predict and control NW shape and diameter without causing much surface damage, it will be extremely useful if a model can be developed based on the study of strain, diffusion and oxidation rate. Although some preliminary exploration based on the stress limited diffusion kinetics has been done recently by Cui *et al.*, we believe more data and detailed model of stress build up and distribution would help paint a clearer picture and facilitate better prediction of NW size and shape.

### *7.2.2 <u>CBCM Measurement for transport modeling</u>*

Understanding carrier transport in nanoscale multi-gate devices is of great importance for the assessment of their performance potential and limits. In this dissertation, we have demonstrated successfully that CBCM can be applied to NW for C-V measurement and extracting the charge and mobility information. On the basis of this novel measurement technique, CBCM test keys can be designed and fabricated with DUT devices of various gate lengths, wire diameters and dielectric thickness, the measurement will generate a pool of data on capacitance and charges with little ambiguity due to uncertainties due to noise, dimension non-uniformity, and defects. Plenty of simulation work using various techniques has been done for NW devices for evaluation of their channel mobility; these experimental results will be extremely useful for validating the simulation models and predictions. The same can be applied to CNT FETs, FinFETs, and other nanoscale devices in the family.

### *7.2.3 <u>Optimization for minimized extrinsic capacitance</u>*

The measurement and simulation results presented in Chapter 5 and 6 show that the extrinsic capacitance for NW devices can be even higher than the intrinsic capacitance which will lead to higher delay in circuit applications. An optimized fabrication process and device structure modification will be required to reduce the extrinsic capacitances more aggressively so that they can be appropriately used in logic and memory circuits.

# APPENDIX: LIST OF PUBLICATIONS

1. **Charge Based Capacitance Measurement Technique for Nano-scale Devices: Accuracy Assessment Based on TCAD Simulations**

Hui Zhao, Subhash C. Rustagi, Fa-Jun Ma, Ganesh S. Samudra, Navab Singh, G.Q. Lo, and Dim-Lee Kwong

*IEEE Transaction of Electron Devices, to be published*

2. **Characterization and Modeling of Sub-Femto Farad Nano-wire Capacitance Using CBCM Technique**

H. Zhao, Raseong Kim, Abhijeet Paul, Mathieu Luisier, Fajun Ma, S.C. Rustagi, G. S. Samudra, N. Singh,G. Q. Lo, Dim-Lee Kwong

*IEEE Electron Device Letters, to be published*

3. **Sub-Femto-Farad Capacitance-Voltage Characteristics of Single Channel Gate-All-Around Nano Wire Transistors for Electrical Characterization of Carrier Transport**

H. Zhao, S. C. Rustagi, N. Singh, F.-J. Ma, G. S. Samudra, K. D. Budhaaraju, S. K. Manhas,C.H. Tung, G. Q. Lo, G. Baccarani, and D. L. Kwong

*International Electron Device Meeting (IEDM), Dec, 2008, San Francisco, CA, accepted for oral presentation.*

*IEDM 2008 Tech. Dig. Pp769-772*

4. **Accuracy Assessment of Charge-Based Capacitance Measurement for Nanoscale MOSFET Devices**

Hui Zhao, Subhash C. Rustagi, Fajun Ma, Ganesh S. Samudra, Navab Singh, G.Q. Lo, Dim-Lee Kwong

*International Conference on Solid State Devices and Materials (SSDM), Tsukuba, Japan, accepted for oral presentation*

*Extended Abstracts of the 2008 SSDM, pp886-887*

5. **Analysis of the Effects of Fringing Electric Field on FinFET Device Performance and Structural Optimization Using 3-D Simulation**

Hui Zhao, Yee-Chia Yeo, Subhash C. Rustagi, and Ganesh Shankar Samudra

*IEEE Transaction of Electron Devices, vol. 55, no. 5, pp1177-1184, 2008*

6. **Simulation of Multiple Gate FinFET Device Gate Capacitance and Performance with Gate Length and Pitch Scaling**

Hui Zhao, Naveen Agrawal, Ramos Javier, Subhash C. Rustagi, M. Jurczak, Yee-Chia Yeo, and Ganesh S. Samudra

*International Conference on Simulation of Semiconductor Process and Devices (SISPAD), Monterey, CA, accepted for oral presentation.*

*SISPAD2008 Tech. Dig., pp252-255, 2006*

7. **Charge Based Capacitance Measurements and Its Application to Transport Characterization In Gate-All-Around Nanowire MOSFETs**

H. Zhao, S. C. Rustagi, G. S. Samudra, F.-J. Ma, N. Singh, G. Q. Lo, D.-L. Kwong

*Communicated to IEEE Transaction of Electron Devices.*