THE DESIGN, FABRICATION AND CHARACTERIZATION OF INDEPENDENT-GATE FINFETS

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

David Michael Fried

May 2004

THE DESIGN, FABRICATION AND CHARACTERIZATION OF INDEPENDENT-GATE FINFETS

David Michael Fried, Ph.D.

Cornell University 2004

The Independent-Gate FinFET is introduced as a novel device structure that combines several innovative aspects of the FinFET and planar double-gate FETs. The IG-FinFET addresses the concerns of scaled CMOS at extremely short channel lengths, by offering the superior short channel control of the double-gate architecture. The IG-FinFET allows for the unique behavioral characteristics of an independent-gate, four-terminal FET. This capability has been demonstrated in planar double-gate architectures, but is intrinsically prohibited by nominal FinFET integration schemes. Finally, the IG-FinFET allows for conventional CMOS manufacturing techniques to be used by leveraging many of the FinFET integration concepts. By introducing relatively few deviations from a standard FinFET fabrication process, the IG-FinFET integration offers the capability of combining three-terminal FinFET devices with four-terminal IG-FinFET devices in one powerful technology for SoC or Analog/RF application, to name only a few.

The IG-FinFET device is examined by device modeling, circuit simulation, testsite design, fabrication and electrical characterization. The results of two-dimensional device simulations are presented, and the effects of process variations are discussed in order to understand the desire for a fully self-aligned double-gate architecture. Circuit design is investigated to demonstrate the capabilities of such a double-gate device. Physical designs are also examined, and the layout penalties of

implementing such a device are discussed in order to understand the requirement of double-gate and independent-gate integration. A test vehicle is designed and presented for the structural integration and fabrication process development necessary for the demonstration and validation of this novel device architecture. The processing and results of several fabrication experiments are presented, with physical and electrical analysis. The integration changes and process modifications suggested by this analysis are discussed and analyzed. Fabricated devices are then electrically and physically characterized. The final set of fabricated devices show excellent agreement with simulated devices, and experimental verification of double-gate device theory. The results of this work provide for a new and novel device architecture with wide ranging technology application, as well as a new fabrication platform with which to study double-gate device theory and further technology integration.

BIOGRAPHICAL SKETCH

David Michael Fried was born on July 5, 1976 to Dr. Peter Marc Fried and Mrs. Wendy Ann Fried in Madison, Wisconsin. David graduated from Madison High School, in Madison, New Jersey in 1994. He received his undergraduate education from Cornell University in Ithaca, New York and graduated in December 1997 with a Bachelor of Science in Electrical Engineering, and December 1998 with a Master of Engineering in Electrical Engineering, also from Cornell University. From January 1999 through 2001, David worked at IBM Microelectronics Division in Burlington, Vermont. In January of 2002, David returned to Cornell University. He received a Master of Science in Electrical and Computer Engineering in August of 2003. He completed his Doctor of Philosophy in Electrical and Computer Engineering with a minor in Applied Physics in 2004. David is currently a member of the Exploratory Devices and Integration Department at International Business Machines Corp. T. J. Watson Research Center in Yorktown Heights, New York.

ACKNOWLEDGEMENTS

I wish to acknowledge the support and guidance of my thesis advisor, Professor Kevin T. Kornegay. His advice, direction and oversight were instrumental in the success of this work. I would also like to thank the members of my special committee, Professor Sandip Tiwari and Professor J. Richard Shealy for their guidance and technical contributions. I would like to add a special thanks to the late Dr. J. Peter Krusius, my Master of Engineering advisor, and a member of my special committee before his passing. His guidance through the early years of my graduate work helped to focus my interests in semiconductor devices and microfabrication.

I am indebted to the Cornell Broadband Communications Research Laboratory (CBCRL) for the support and funding of this research through a Graduate Research Assistantship. This work was performed, in part, at the Cornell Nano-Scale Science & Technology Facility (a member of the National Nanofabrication Users Network) which is supported by the National Science Foundation under Grant ECS-9731293, its users, Cornell University and Industrial Affiliates. I would like to strongly thank the staff at the Cornell Nanofabrication Facility (CNF), specifically Meredith Metzler, Garry Bordonaro, Mike Skvarla, Phil Infante, Alan Bleier, Jim Clair, Rob Ilic, and Daron Westley for all of their training, guidance, and assistance. This work also made use of the Cornell Center for Materials Research Shared Experimental Facilities, supported through the National Science Foundation Materials Research Science and Engineering Centers program (DMR-0079992). The LEO 1550 SEM was originally funded by the Keck Foundation, with additional support from the Cornell Nanobiotechnology Center (STC program, NSF award # ECS-9876771)

Present and former members of the CBCRL, the Advanced Integrated Micro-Systems (AIMS) research group, and Professor Kornegay's research group are also

acknowledged. In particular, I would like to thank Dr. Jon S. Duster for his leadership, technical assistance and attitude modulation. I would like to thank other double-gate enthusiasts from Cornell, including Uygar Avci, Arvind Kumar, Helena Silva, and Gen Pei, for fruitful arguments over the best embodiment of the double-gate architecture. I would like to thank other CNF users, including Jeremy Wahl, Udayan Ganguly, and Ali Gokirmak for assistance and moral support in the lab. I would like to thank students Ian Rippke, Dr. Sean Welch, Drew Guckenberger, Daniel Kucharski, Kyle Maurice, Franklin Baez, Yanxin Wang, Conan Zhan, Rich Gabrielson, Brian Welch, Dr. Mihaela Balseanu, Dr. Ce Li, Dr. Eskinder Hailu, Dr. Andrew Atwell, Dr. Paul Ampadu, Javier Alvarado, and Jami Meteer for their support, friendship, and partnership in bowling, sailing, basketball, and other diversions from research.

I would like to acknowledge the support I received from International Business Machines Corp. and specifically Timothy Macnee, Dr. Meikei Ieong, Dr. T. C. Chen, Dr. Ghavam Shahidi for securing my position in the Special Studies Plan, and for supporting and encouraging my research. I wish to acknowledge the technical assistance I received from Dr. Jakub Kedzierski, and other FinFET researchers at IBM. I wish to pay particular thanks to Dr. Edward Nowak who has had a profound effect on my personal, professional and technical development. His support, teaching, advice, guidance and friendship have contributed to all of my successes, and my addiction to caffeine.

I would like to recognize the support, guidance, advice and love that I received continuously from my wife, Marina. Without her, none of this work would have begun, progressed or come to completion. I have been lucky to have the support of my entire family, my sister Lisa, my brother Stephen, my parents Wendy and Peter, as well as my in-laws, John, Silvia, Tricia and Evi and extended family members.

This work is dedicated to my father, Dr. Peter Marc Fried, who has shaped much of the way I think, instilled in me confidence and technical desire, and unknowingly coined many of the phrases by which I live my life.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1: Introduction

1.1 Perspective

Since the first integrated circuits were fabricated, the ambitions of technology development have remained constant: increased functionality, enhanced performance, and decreased cost. The industry focus on Silicon Metal-Oxide-Semiconductor Field Effect Transistors (MOSFETs) for the past 50 years has been mainly due to the simplicity of manufacture, inherent scalability and high levels of integration possible. This single-minded focus has led to state-of-the-art Silicon Complementary-MOS (CMOS) front-end technology with gate lengths of 50 nm, dielectric thicknesses of 12 Å, and aggressively scaled memory circuits such as $0.6um^2$ SRAM and $0.11um^2$ embedded DRAM cells [1]. The development to this level of technology has largely been done with the same materials and structures as the earliest MOSFETs: a metal (Aluminum, Tungsten) or semi-metal (Doped Polysilicon) gate electrode on an insulating (Oxide, Nitride, Oxynitride) gate dielectric above a crystalline or poly-crystalline semiconductor (Silicon, Germanium) body. Implanted or diffused ions (Boron, Indium, Phosphorus, Arsenic, Antimony) have been used to alter the carrier statistics in source, drain, gate and body regions of the device. Fabrication techniques have advanced significantly to allow the integration of smaller features, thinner films and more controlled dopant profiles. However, very few significant changes were made to the actual structure of the MOSFET until very recently.

As gate lengths have scaled below 250 nm for increased performance and integration, several physical effects have begun to challenge device designers to reach scaling performance targets. Many of these physical effects fall under the broad category of Short Channel Effects (SCE). In general, SCE arises from several geometrical effects that decrease how effectively the gate electrode is able to modulate the drain current of the FET. A major effect of SCE is Drain Induced Barrier Lowering (DIBL), whereby the drain, as gate lengths are scaled shorter, moves closer to the source-to-channel potential barrier. High electric fields from the drain can lower that barrier that is supposedly only controlled by the gate. This effect can degrade the devices' Subthreshold Slope and cause changes in the Threshold Voltage (V_T) as a function of drain bias.

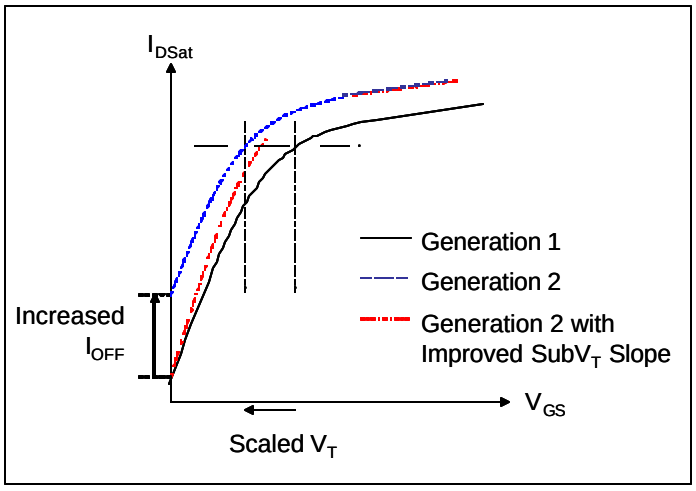

Figure 1. Idealized I_D vs. V_{GS} showing why Subthreshold slope improvement is required when scaling V_T

In order to increase the gate control of the source-to-channel barrier, gate dielectrics must be made thinner, effectively increasing the gate capacitance and gate-to-barrier coupling relative to drain-to-barrier coupling. Thinner gate dielectrics allow more tunnel current to pass between the gate electrode and the body of the

device. This gate current results in excess power dissipation and can result in damage to the transistor. These thinner gate dielectrics require that supply voltages are lowered to prevent breakdown and damage to the dielectric. To scale the supply voltage, V_T must also be scaled. Without improvements to Subthreshold Slope, lowering V_T will result in increased off-current as seen in Figure 1. Increased off-current adds to gate current in increasing the chip-level static power dissipation.

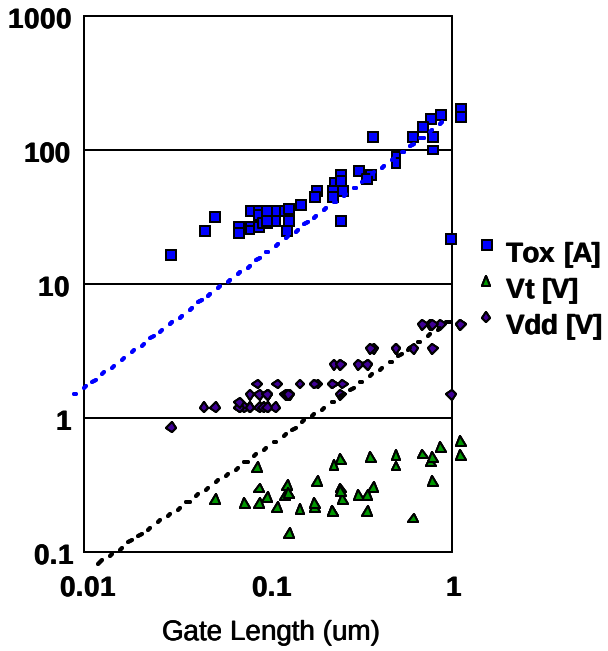

Figure 2. Industry data showing T_{OX}, V_T and V_{DD} scaling trends below 1micron gate length [2]

Equipment and fabrication innovations have allowed the device designers to continue to scale the channel length of transistors, along with the gate dielectric thickness well beyond expectations. However, this scaling has come at the expense of power dissipation. For technologies past the 90 nm lithography node, limitations on gate current have begun to limit the gate dielectric scaling as seen in Figure 2, and SCE has becoming more difficult to overcome with traditional scaling. Several

material innovations have been suggested to combat SCE, such as high-K gate dielectrics. These materials, such as HfO_2 and ZrO_2, may provide increased gate capacitance with thicker films that present a larger tunneling barrier, and therefore lower gate current [3]. While much research on these materials is ongoing, many material and integration challenges have so far prevented their introduction into high performance Silicon logic technologies.

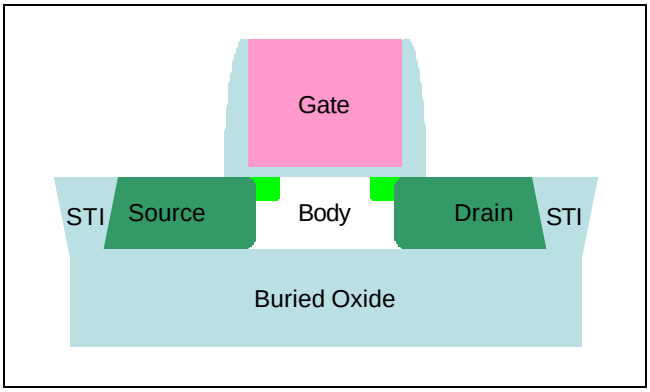

Figure 3. Structural cross-section of Silicon on Insulator (SOI) Device

The first major structural innovation to address these concerns was Silicon-on-Insulator (SOI) [4]. By fabricating MOSFETs in a thin single-crystal Silicon film above a SiO_2 film as shown in Figure 3, device designers were able to decrease the size of the drain to body junction, and decrease the fields that were able to couple to the source-to-channel barrier. This structural change eased some of the constraints on gate dielectric thickness. The other major benefit of SOI is that the source and drain have their bottom surface on oxide, instead of in silicon. The lower dielectric constant of oxide equates into decreased source and drain junction capacitance. This improvement is manifested in AC performance, in the charging and discharging of circuit nodes during switching. One major disadvantage of the first SOI devices was that they were not fully-depleted. The silicon film that the devices were fabricated in was thick enough such that there still existed a quasi-neutral region in the body of the

device. Since the device, known as Partially-Depleted SOI (PD-SOI) is isolated by the Buried Oxide (BOX), any majority charge that builds up in the body due to impact ionization is held contained in this quasi-neutral region as shown in Figure 4. This charge can affect the device performance by altering the V_T, similar to a well bias in bulk CMOS. Since the amount of built up charge in the body was dependant on the transient state of the device, the V_T could fluctuate during operation. This became known as the Floating Body Effect [5].

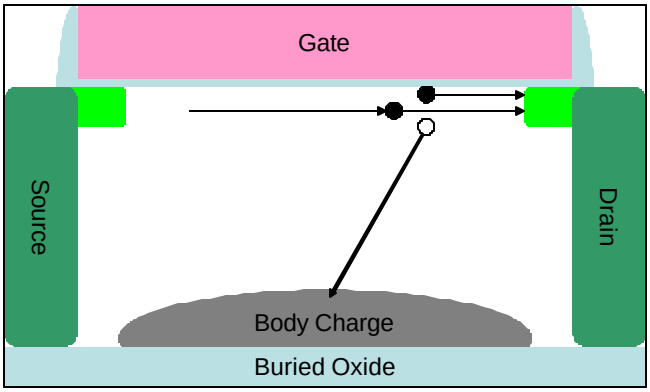

Figure 4. Cross section of active device area showing impact ionization and the body effect

By making the SOI film thin enough so that at zero bias the body was fully depleted of majority carriers, device designers were able to eliminate the Body Effect. Fully Depleted SOI (FD-SOI) allows for increased performance as a result of eliminating the majority carriers in the body, and by further shrinking the drain to body junction [6]. However, FD-SOI integration is significantly more difficult than PD-SOI. The thin SOI film significantly increases the parasitic source/drain resistance. And, because it is very difficult to produce an extremely constant dopant concentration in the small volume of the device body, V_T must be adjusted by modifying the SOI thickness, the gate dielectric thickness and the gate electrode work-function.

1.2 Motivation

Double-Gate CMOS (DG-CMOS) represents the next geometrical structural innovation to control SCE in Silicon FETs for high performance logic [7, 8]. By placing gate electrodes on both sides of a thin silicon body as in Figure 5, several device advantages can be achieved. First, for a given gate electrode work-function, gate dielectric thickness, and body doping, the DG-CMOS device will be fully-depleted with a thicker body than its FD-SOI equivalent. This allows for increased performance from full depletion without the severe source/drain resistance penalty. Second, the presence of the additional gate suppresses the fringing drain electric fields that cause SCE and specifically DIBL [9]. Third, if both sides of the device (top and bottom) are used for current conduction, it is possible to significantly increase current density per layout area. Finally, due to several physical mechanisms, such as the fact that double-gate devices operate at significantly lower transverse electric field, mobility is often enhanced [10].

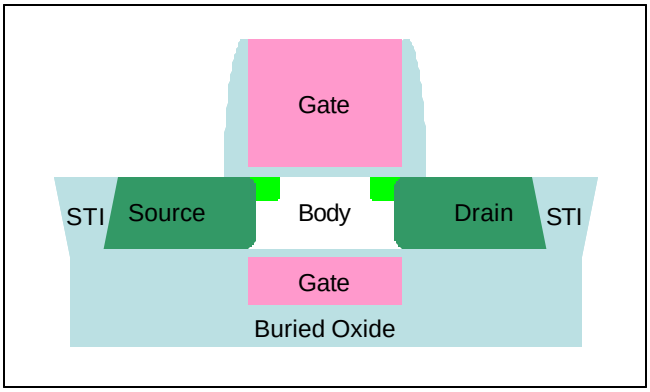

Figure 5. Structural cross-section of planar double gate FET

The fabrication and integration of DG-CMOS poses its own serious fabrication challenges. The first double-gate devices were of the planar variety, with the back gate buried in the BOX layer. Integration techniques for these structures were

extremely complex, often involving wafer-bonding and other processes not typically used in CMOS manufacturing [11-13]. The two main challenges to fabrication of planar DG-CMOS are the alignment of the top and bottom gates and the fabrication of the back-gate dielectric. The two gates are typically etched individually at the minimum lithographic dimension, and therefore ensuring perfect gate-to-gate alignment is quite difficult. If the gates are even slightly misaligned, the performance enhancement may be degraded due to additional overlap capacitance and other effects [14]. Keeping the back-gate dielectric free of contamination during these complex integration processes is also extremely difficult.

The FinFET has gained popularity recently as a potentially simpler device to fabricate that still leverages the benefits of DG-CMOS for performance [15-23]. A FinFET is fabricated by etching a thin "fin" of Silicon through the SOI layer, stopping on the BOX layer. Gate dielectric is grown on the surfaces of the fin, and a gate electrode material is deposited, patterned and etched. This fabrication process is illustrated in Figure 6. Since the gate dielectrics are grown simultaneously, the contamination risk is decreased. Since the gate electrodes are patterned with a single lithographic shape and exposure, gate misalignment is also decreased. All of the processing and materials used to fabricate the FinFET are standard to conventional CMOS process technology, due to the inherent top-down nature of the FinFET integration scheme. The disadvantages of a FinFET are the fact that the body thickness of the device is now controlled by an etching operation. Therefore there may be more variation in the body thickness than in a conventional SOI device. Since the body thickness controls V_T in a fully-depleted device, this effect is quite concerning. Also, the current in a FinFET is carried on the sidewalls of this etched fin. Surface roughness effects may be more substantial in this device that conventional SOI devices where the current is carried on a polished surface.

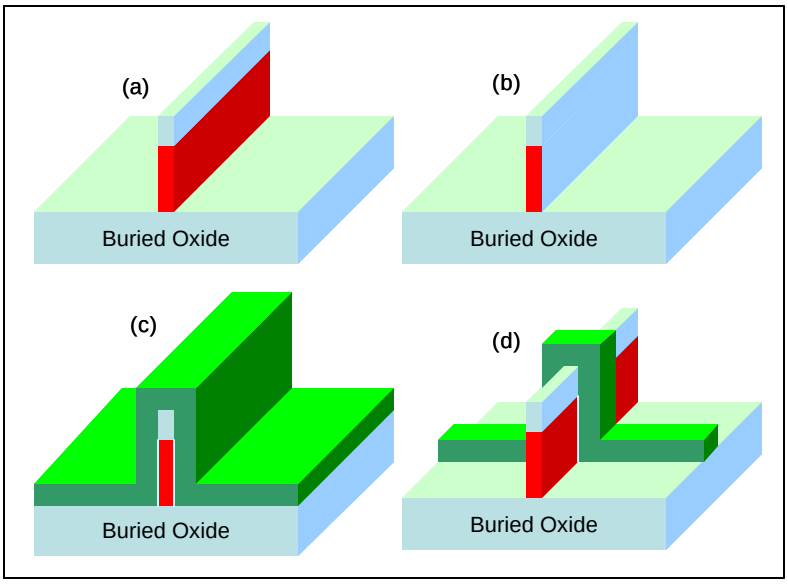

Figure 6. Fabrication process of a FinFET (a) The fin is etched in SOI. (b) The sidewalls are thermally oxidized. (c) Polysilicon is deposited. (d) The gate electrode is etched.

Comparing planar DG-CMOS to FinFET CMOS becomes even more difficult when circuit design and layout considerations are included. Many innovative circuit design techniques utilize a well contact to bias the well in bulk CMOS and PD-SOI. This well bias can be used to adjust the V_T of the device, and can be used in certain analog circuit as another input in addition to the gate. However, in a fully-depleted SOI device, this well bias is useless, as the majority carriers are already depleted away from the body, and a contact to the fully depleted semiconductor would not provide means for adjusting the potential. Therefore, in FD-SOI devices, the body contact terminal has been eliminated, and the device becomes a three terminal FET. DG-CMOS devices are slightly different. If the gates are electrically connected, the device behaves like a three-terminal high-performance FD-SOI device. If the gates are electrically isolated and individually accessible, then the device behaves again like a four-terminal FET, but different from a bulk FET with a well contact. This four terminal DG-CMOS device can be used in a circuit design style similar to active well

biasing, and in novel analog circuit design techniques. The exact behavior of this device will be discussed later.

1.3 New and Original Contributions

A technology that would provide circuit designers with both true Double-Gate FETs ("DG" - gates electrically connected) and Independent-Gate FETs ("IG" - gates electrically isolated and accessible) could be extremely powerful for System-on-Chip (SoC) integration, mixed Analog/Digital integration and novel low-power circuit design. Most planar DG-CMOS devices are fabricated using schemes that naturally result in electrically isolated gates. They are individually accessible by design. Tying these gates together requires additional layout area outside the active region to fabricate a gate-to-gate via, not to mention a process to fabricate these vias. If the device has a wide active region, there must be a fabrication technique for ensuring low enough resistance through the long run of back gate electrode. If the process exists to contact each of the gates in a planar DG-CMOS technology, then the designer could choose DG or IG devices simply with layout changes. The FinFET has inherently connected gates as the gate electrode wraps up and over the fin. No additional layout area is required to connect the gate, and, as such, the FinFET is slightly more layout efficient for wide area DG devices. Since the FinFET is fabricated with top-down fabrication processing, a conventional Silicide process could be used to lower gate resistance, and would affect both "front-gate" and "back-gate" resistance. This inherent gate connection is quite useful for the DG devices, but makes IG devices impossible.

The focus of this work is the first demonstration of a new double-gate architecture, the Independent-Gate FinFET. Circuit design and graphical layout implications have been investigated, and a comprehensive testsite design has been

completed. A fabrication scheme for an Independent-Gate CMOS device in a FinFET based technology has been developed and integrated. This IG-FinFET integration scheme was designed with the intent of being integrated with conventional FinFETs to provide a powerful combination of high-performance devices and novel four-terminal devices. In the interest of integration ability, much of the conventional FinFET process flow must be maintained throughout the process flow for the Independent-Gate FinFET. Several integrated process runs have been completed to explore the possibility of fabricating this new device. IG-FinFET processes have been developed to improve performance and yield. Finally, electrical characterization has been performed on completed devices to show the success of process development. Final devices show excellent characteristics and agreement with simulation data.

Chapter 2: Double-Gate Physics and Device Models

2.1 Overview

In order to better understand the design of Independent-Gate FinFETs, the parameters of interest and the various aspects of the fabrication process, a brief review and analysis of Double-Gate MOSFETs (DGFETs) and the physics that govern their behavior is necessary. A great deal of literature is available regarding the device physics involved in double-gate devices, specifically those used in double-gate mode (both gates tied together). Less literature has been published regarding the use of these structures in independent-gate mode. This chapter first presents the central physics that control the potential profile in the body of a double-gate structure, including the device design parameters of interest. The differences introduced into the salient device physics when the gates are decoupled are discussed. Finally, this chapter concludes with an examination of the device parameters that are affected by fabrication processes, and the effect of these tolerances on the device. This analysis leads to a list of advantages and disadvantages associated with different independent double gate device structures.

2.2 Definition of Parameters

Since many of the parameters for this device are replicated on both sides of the body, a few definitions must be made. Table 1 lists several important device parameters. Many of them are also indicated graphically on Figure 7.

Table 1. Device parameters and definitions

Parameter	*Units*	*Description*
L_{GATE1}	nm	Physical Top/Front Gate Electrode Length
L_{GATE2}	nm	Physical Bottom/Back Gate Electrode Length
L_{EFF1}	nm	Top Gate Effective Gate Length
L_{EFF2}	nm	Bottom Gate Effective Gate Length
T_{SI}	nm	Physical Body Thickness
N_{BODY}	cm^{-3}	Dopant Concentration in Body
$N_{S/D}$	cm^{-3}	Dopant Concentration in Source/Drain
T_{OX1}	nm	Top Oxide Thickness
e_{OX1}	n/a	Top Oxide Relative Permativity
T_{OX2}	nm	Bottom Oxide Thickness
e_{OX2}	n/a	Bottom Oxide Relative Permativity
F_{MS1}	eV	Workfunction Difference between Top Gate and Body
F_{MS2}	eV	Workfunction Difference between Bottom Gate and Body

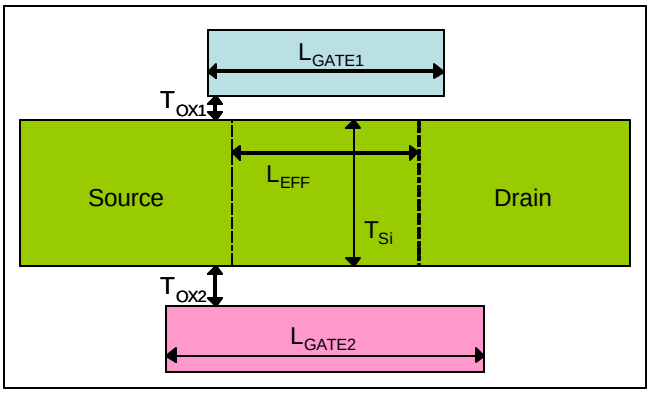

Figure 7. Structural parameters of interest in a double-gate FET

2.3 Double-Gate Capacitor Simulation

As with conventional single-gate devices, an intuitive understanding of the functioning of the composite transistor begins with a description of the device stack in the form of a simple capacitor without the source and drain. The simplest form of DGFET is one with identical gates electrodes and dielectrics, and having fully constant doping through the body in the dimension perpendicular to the gates. Since this work deals with the IG-FinFET, fabricated by growing both gate oxides simultaneously and by patterning and etching both gates simultaneously, this simplification is applicable. In this embodiment, the following simplifications can be made:

$$T_{Ox1} = T_{Ox2} = T_{Ox}$$
$$e_{Ox1} = e_{Ox2} = e_{Ox}$$
$$\Phi_{MS1} = \Phi_{MS2} = \Phi_{MS}$$
$$L_{Gate1} = L_{Gate2} = L_{Gate}$$

Equation 1. Simplications made to parameters for symmetric double gate FET

In a DGFET fabricated with doped polysilicon gate electrodes on a Silicon body that is either thick or heavily doped, there exists enough charge on either side of the body in the form of depleted dopant atoms to mirror the gate charge. Due to this abundance of charge near the dielectric interfaces, the depletion regions are small relative to the thickness of the body. As a result, the potential profile reaches equilibrium in the middle of the body, as shown in Figure 8. Since the depletion depth is small with respect to the body thickness, there exists a zero-field region in the middle of the body. This thick-body DGFET behaves like two PD-SOI devices that share a well region.

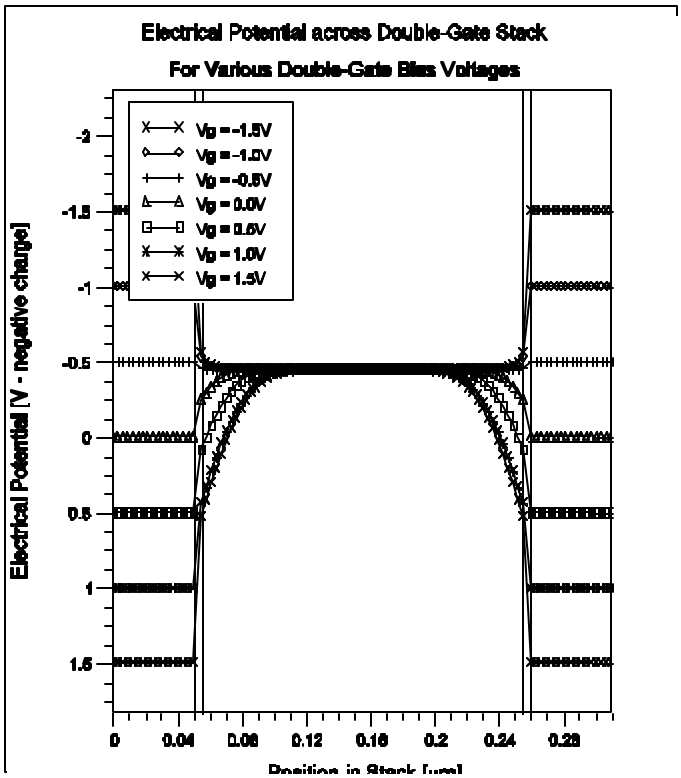

Figure 8. Simulated potential profile of a thick-body double-gate FET at multiple double-gate bias voltages ranging from accumulation through depletion and into inversion

This can also be proven analytically by integrating charge outward from the center of the symmetric device structure. This yields an expression for the band bending across the body of the device.

$$\frac{q(y - y_o)}{2kT} = -\ln\left[\cos\left(\sqrt{\frac{q^2 n_i}{2e_{Si}kT}e^{\frac{qy_o}{2kT}}}\,x\right)\right]$$

Equation 2. Solution of potential profile across the body of a symmetric double-gate device

where $?_o$ is the potential in the center of the body of the device [24]. If the body is significantly thinner or more lightly doped, the depletion regions, now larger relative to the body thickness, may contact each other in the middle of the body, then

the potential profile does not reach equilibrium at any point in the body, as shown in Figure 9. The expression for the potential in the body still applies, as long as the device remains symmetric.

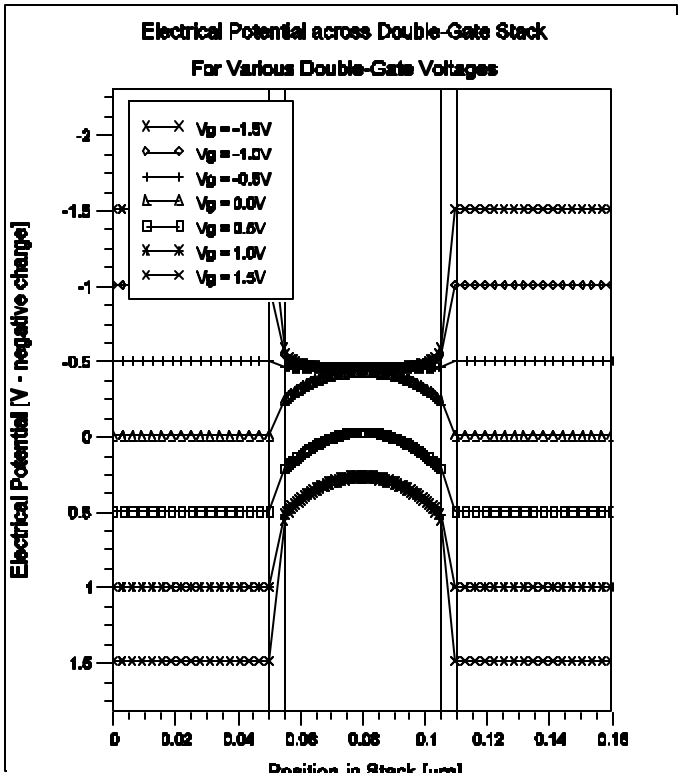

Figure 9. Simulated potential profile of a thin-body (fully-depleted) double-gate FET at multiple double-gate bias voltages ranging from accumulation through depletion and into inversion

The body thickness at which the depletion regions contact one another depends on the body doping level and the gate stack. Figure 10 shows the depletion depth at the onset of strong inversion and the depletion depths for several different gate stacks as a function of body doping.

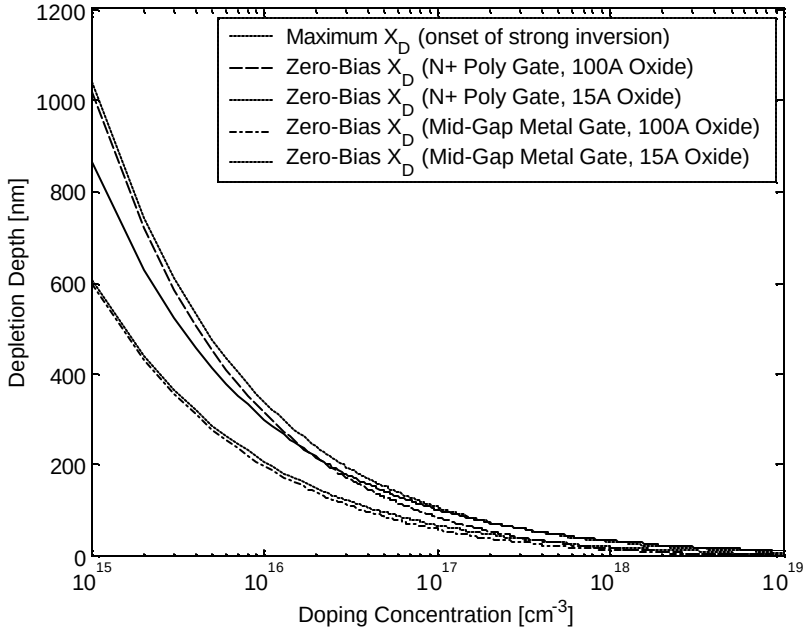

Figure 10. Depletion depth vs. doping level for different gate stacks

If the body thickness is smaller than the sum of the depletion depths of both gates, then the DGFET will be fully-depleted at zero-bias. In this case, changes in the potential profile resulting from activity on one gate can substantially affect the potential profile on the opposite side of the body. From Figure 10, it is clear why device designers have been forced to increase the body doping to account for thinner gate dielectrics, to prevent devices from entering inversion at zero bias. This increase in body doping has lead to two problems in device design. First, the increase in doping can increase the scattering in the body, thereby limiting carrier mobility. Also, since the active device volume in state of the art devices is so small (L_{EFF} x W_{EFF} x t_{DEP}), it has become quite difficult to accurately control the doping concentration between many devices. This is known as the Random Dopant Fluctuation Effect. Because of these two problems, a device with an undoped body would be preferable. In the case of a symmetric DGFET with an undoped body, no field exists between the

gates and the body, and therefore the potential is flat across the gate stack at zero-bias, as seen in Figure 11.

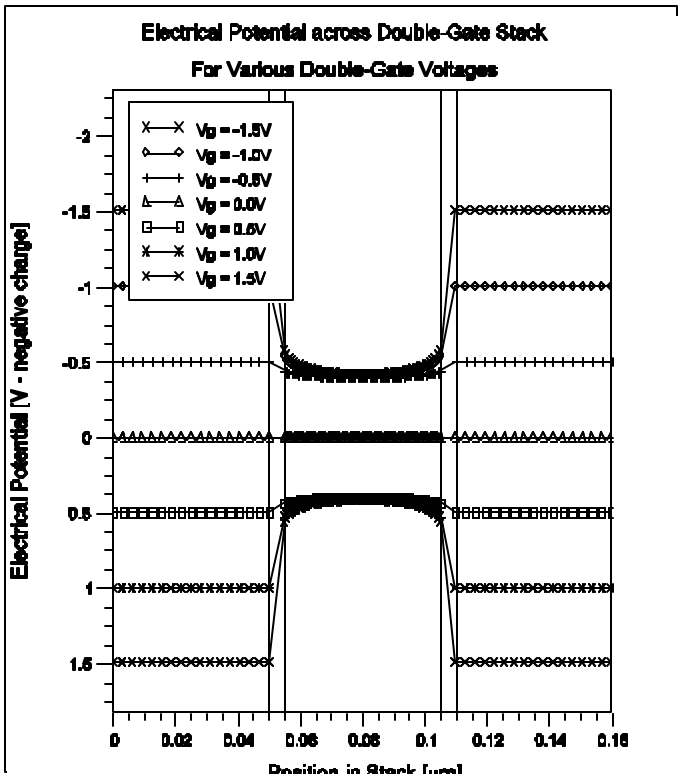

Figure 11. Potential profile of an undoped-body double-gate FET

This device would be strongly inverted at zero-bias due to the relative positions of the Fermi levels in the Source/Drain and the channel, so further gate workfunction engineering would be required for CMOS integration.

In any of these double-gate devices, the depletion regions on both sides of the device expand and contract as a function of applied gate bias. As in conventional planar CMOS devices, as the gate bias is increased, the depletion region extends deeper into the body, until inversion conditions exist, after which the depletion region does not extend appreciably. In fully-depleted SOI devices, this depletion region extension can only proceed until it reaches the interface between the SOI and the buried oxide. At that point, inversion must occur in order to provide charge in the

channel to mirror the charge placed on the gate electrode. In symmetric double-gate devices, this extension can only proceed until it reaches the depletion region created by the opposite gate. Since the gates move in concert in a true double-gate device, the condition where the depletion regions meet is controlled totally by the device geometry and doping parameters.

The analytical solution for the double gate device becomes quite difficult if the device is asymmetric, either in doping, oxide thickness or gate workfunction. The solution of this profile involves the definition of a point in the body x_o used as a boundary condition. E_o is defined to be the field at that point, because in certain configurations there is no zero field region or point in the body [24].

$$y(x) = \frac{2kT}{q} \ln\left\{ \frac{\sqrt{\frac{e_{Si}}{2kTn_i}}\mathrm{E}_o}{\sinh\left[\frac{q\mathrm{E}_o(x - x_o)}{2kT}\right]} \right\}$$

Equation 3. Solution of potential profile across the body of an asymmetric double-gate device

While this solution is quite complex, and developed to solve for the potential profile for an asymmetrically fabricated double-gate device, it is applicable for symmetric independent-gate devices. In an independent double-gate device, the ability to vary both gate biases mimics the variation in workfunction that drives the solution above.

2.4 Independent-Gate Capacitor Simulation

In an independent double gate device, the depletion regions are not necessarily identical. In independent-gate devices, the opposite depletion region is

controlled by the opposite gate, somewhat independently from the front gate. In the case of the thick-body or heavily-doped IG-FET, the inversion condition is reached before the two depletion regions get close enough to one another to affect device attributes. Therefore, the bias on one gate has no effect on the condition at which the opposite channel enters inversion. However, in a thin-body or lightly doped body independent-gate device, this is not true. The depletion region of the opposite gate can be pushed well into the range where the front depletion region would ordinarily extend. In this case, the back gate can be used to control the condition where the front depletion region can no longer proceed, and inversion begins. In a device with an extremely thin and lightly doped body, the depletion regions meet even at zero bias. In this case, the back gate bias can control the potential profile across the entire body of the device, including very close to the front gate interface, thereby altering the conditions where inversion is reached on the front channel. This effect is exhibited in Figure 12, as the potential profile is altered significantly across the entire body of the double-gate stack, solely as a function of the back gate bias.

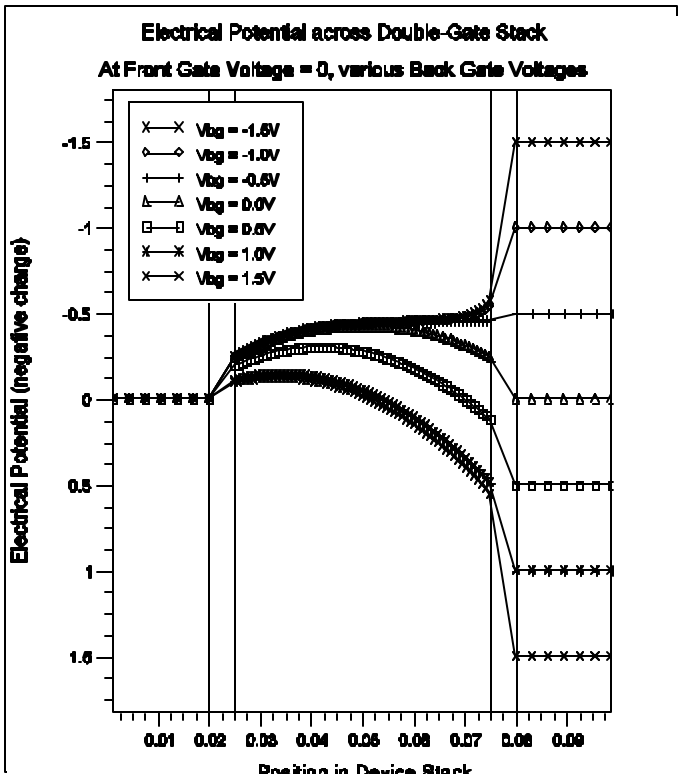

Figure 12. Simulated electrical potential profile across a double-gate stack with 50nm thick body with 1e17 cm^{-3} Boron doping and symmetric 50A gate oxides, as the front gate is held at zero potential and the back gate is biased from -1.5V to 1.5V

This change in electrical potential also causes a change in the electron concentration on both interfaces of the double-gate stack as seen in Figure 13 under the same bias conditions. In the lower range of back gate bias ($V_{BG} < 0$), the electron concentration on the front interface is nearly constant, but the potential profile near the front interface changes slightly.

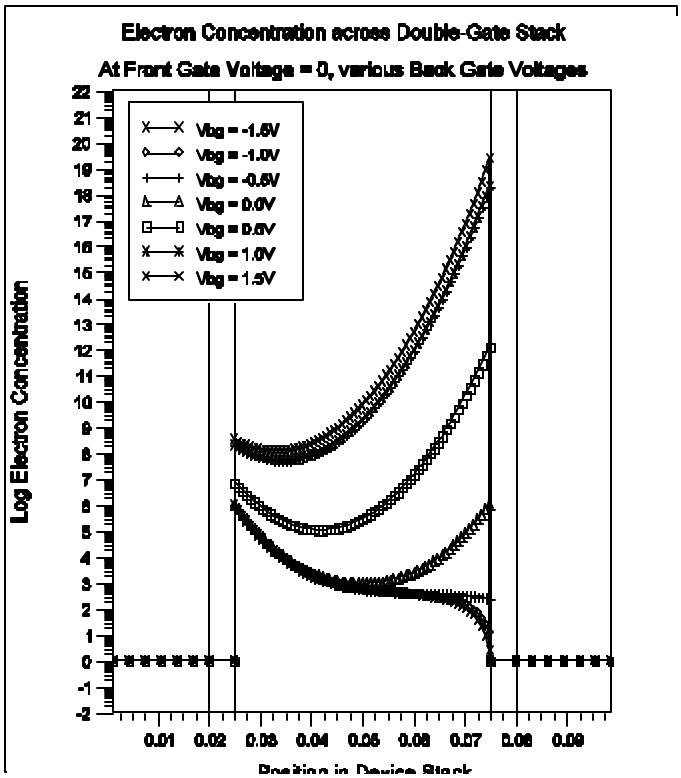

Figure 13. Simulated electron concentration across a double-gate stack with 50nm thick body with 1e17 cm-3 Boron doping and symmetric 50A gate oxides, as the front gate is held at zero potential and the back gate is biased from -1.5V to 1.5V

2.5 Independent Double-Gate Transistor Simulation

The addition of Source/Drain regions to the capacitor structures explored in the previous sections has a similar effect as in single gate CMOS and SOI devices. With source and drain regions present, the small variation in electron concentration at the front gate interface will result in small drain current density variation. However, the change in the potential profile will alter the condition where inversion occurs for the front interface as a function of the front gate voltage, and therefore alter the threshold voltage. Once the back gate bias proceeds to higher values ($V_{BG} > 0$), the back interface enters inversion, providing a channel from source to drain. Under these conditions, the change in potential profile induced at the front interface has little effect

on the current density from source to drain as the back interface is providing the dominant current path between source and drain. This is demonstrated using device simulations similar to the ones used in the capacitor analysis. In these simulations, an idealized Independent-Double gate device is used as shown in Figure 14.

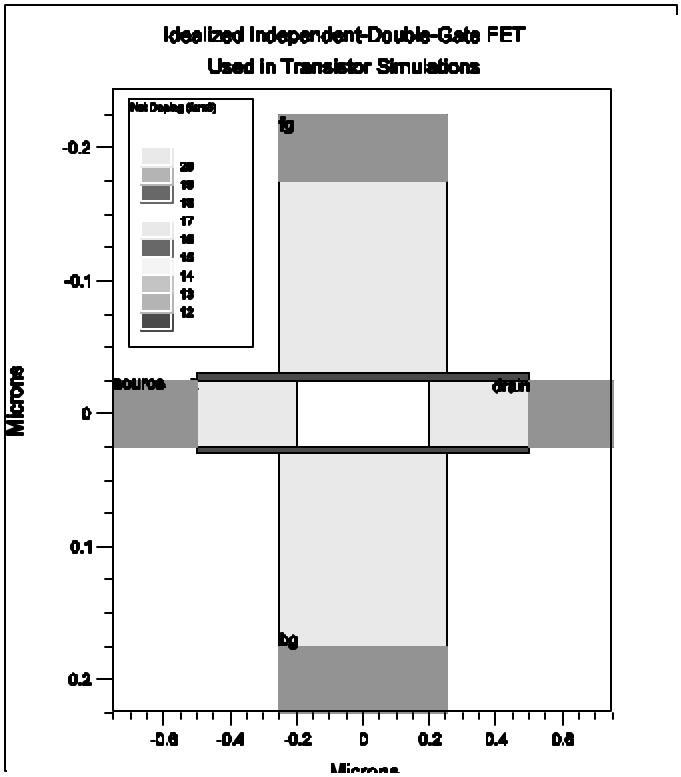

Figure 14. Idealized Independent Gate Device Structure used in transistor simulations with T_{FIN} = 50nm and L_{GATE} = 500nm with 50nm symmetric S/D Overlap, 1e17 Boron body doping, and 1e20 Arsenic doping in the Source/Drain and both gates

This device was simulated through a large range of front and back gate bias voltages, in both linear (V_{DS} = 50mV) and saturated (V_{DS} = 1.5V) conditions.

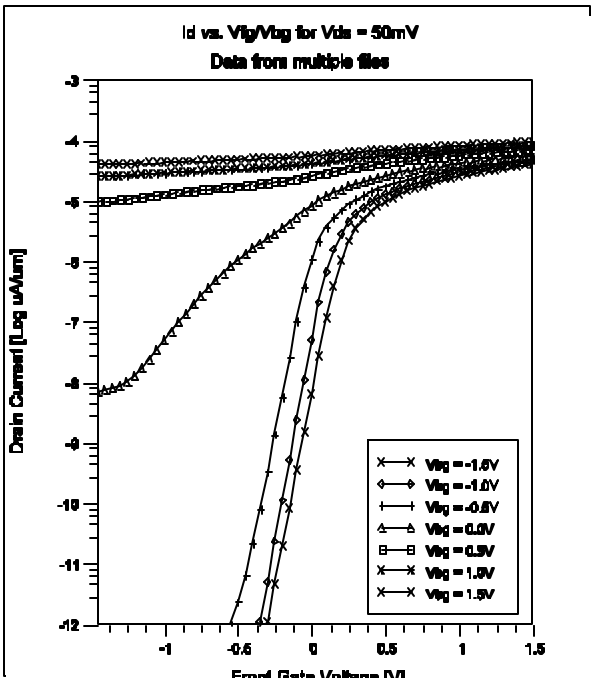

Figure 15. Linear (V_{DS} = 50mV) Drain Current as a function of front and back gate bias voltages

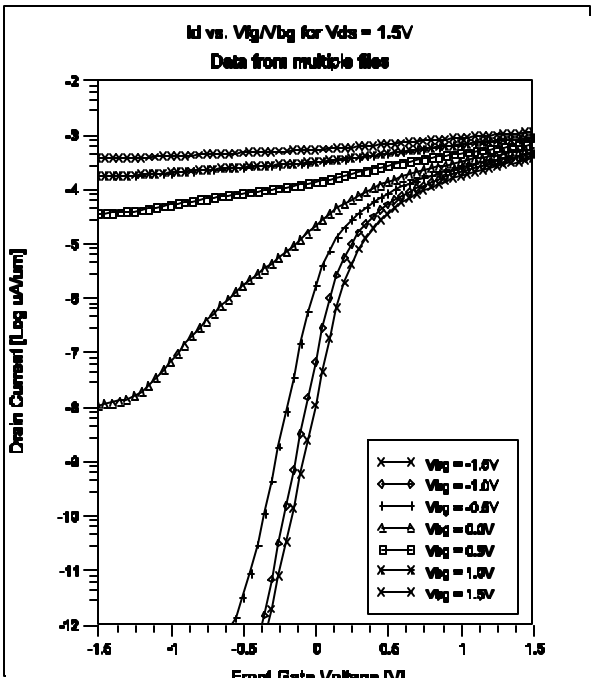

Figure 16. Saturation (V_{DS} = 1.5V) Drain Current as a function of front and back gate bias voltages

As seen in Figure 15 and Figure 16, the changes in potential profile across the body of the device, as discussed in the previous section, have the expected effects on drain current. At low back gate bias, when the effect of the back gate is simply a change in the potential profile across the body, the effect on the device is to modify the threshold voltage by changing the source to channel barrier height. This effect is limited to the range when the back gate bias does not cause inversion conditions in the back interface. Once this inversion condition is reached in the back channel, the aggregate device is on, regardless of the front gate bias. The front gate remains active in control of the front channel, however, this current is comparable, in the symmetric device, to the back channel current, and therefore a linear addition to the aggregate device current.

2.6 Effects of Gate Misalignment

Many different structures have been envisioned to enable this type of independent double gate behavior. The simplest of these structures, often referred to as a Ground-Plane Device , utilizes a non-self aligned, blanket buried electrode to serve as the back gate as seen in Figure 17. This structure is less difficult to fabricate that many other double-gate structures due to the fact that the gates require no alignment scheme and the back gate dielectric interfaces may be protected during the entire fabrication process. From a DC electrical standpoint, this device behaves nearly identically to the fully self-aligned independent double gate device. However, in circuit operation, this device has one severe drawback. The excessive back gate overlap of the source and drain causes large overlap capacitance. This capacitance is directly coupled to the circuit performance, as the source and drain nodes must charge and discharge during switching conditions. There may exist a design window where a thick enough back gate dielectric may provide ample control of the potential profile in

the body without adding unacceptable overlap capacitance, but this requires detailed analysis.

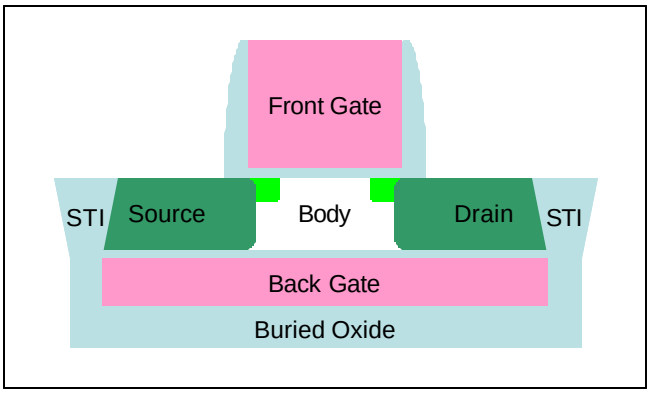

Figure 17. Cross-sectional structure of the Ground-Plane FET

Because of this drawback, much research has been directed at the fabrication of double-gate devices, both independent-gate and true double-gate, where the back gate length dimension is reduced to provide ample control of the potential profile in the body of the FET and still maintain the junction capacitance and overlap capacitance advantages of planar SOI. The optimal structure would provide identical sized gate electrodes, perfectly aligned to one another, and also perfectly aligned to the source and drain. Due to the lithographic limits of fabrication technology, "perfect alignment" of gates and implant junctions requires self-alignment, as in the source and drain in planar CMOS.

Several of the planar double-gate embodiments require two separate lithographic steps to pattern the two gates. This inevitably introduces statistical variation in the gate-to-gate alignment of these devices. Based on the fabrication processes used to produce the source/drain regions of these devices, the implant junctions are typically self aligned to one gate, but not the other. Top down implants would result in source/drain regions aligned to the top gate, but not the bottom gate. Therefore, the gate-to-gate misalignment directly translates to a misalignment of one

gate to the source/drain junctions. This gate misalignment has a serious effect on the performance of the device. If the gate is misaligned enough in either direction (towards the source, or towards the drain), there will exist a region on the opposite side of the back gate where the body of the device is not double-gated. Here, the advantage of the double-gate device is lost. If this un-gated region is on the drain side, the back gate is not preventing the drain fields from fringing into the channel and causing threshold voltage roll-off. This results in increased DIBL for this device. If the un-gated region is on the source side, the back gate is not assisting the front gate in controlling the source-to-channel barrier. This results in a change in the absolute threshold voltage of the device. In this case however, the additional drain overlap will help to control the fringing drain fields better, thereby reducing the DIBL. Figure 18 shows the saturated and linear I_D vs. V_{FG} curves for V_{BG} = 0.0V. The DIBL increase and reduction are clearly apparent for a 500nm gate misaligned by 100nm in either direction. This amount of misalignment is consistent with advanced lithographic technology where overlay or alignment tolerance may actually exceed 20% of the minimum feature size. Although the effect of this misalignment may be acceptable in certain circumstances, such as the DIBL reduction caused by drain-side misalignment, the circuit level effect is clearly unacceptable. This gate-to-gate misalignment is caused by statistical variations in the lithographic processes. Therefore, the misalignment is unpredictable, and the major effect is the variation between the cases shown in Figure 18, not the absolute conditions shown in one case. This variation would make circuit design impossible using these devices. It is for this reason that an optimal double-gate device must have both gates defined with a single lithographic process.

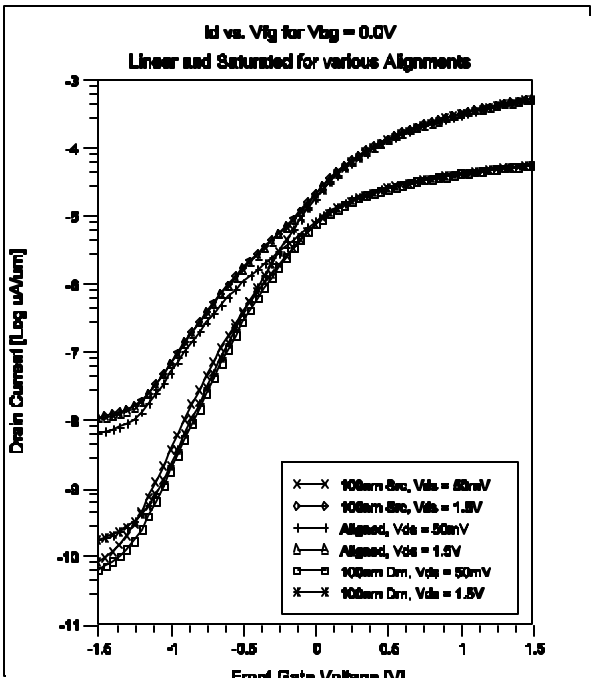

Figure 18. Effect of gate misalignment on independent double-gate device for V_{BG} = 0.0V

2.7 Effects of Fin Thickness on IG Behavior

Much work has been conducted to determine the effects of fin thickness on double gate devices in double gate mode. Fin thickness affects several aspects of device behavior in double gate mode, including carrier concentrations, threshold voltage, capacitances and, in some cases, carrier mobility. In independent gate mode, the fin thickness has one additional interesting effect. Changes in fin thickness affect the way the back gate can control the front channel inversion population. Intuitively, as the fin thickness is decreased the back gate should be more tightly coupled to the front gate due to the decreased absolute amount of fixed charge in the body on which to terminate field. Results of simulations similar to the previous ones, with varied fin thicknesses, show that this intuitive effect is somewhat correct. However, it is a bit difficult to extract the threshold voltage control via the back gate from the DIBL and

other geometrical short channel effects that also change substantially as a function of fin thickness.

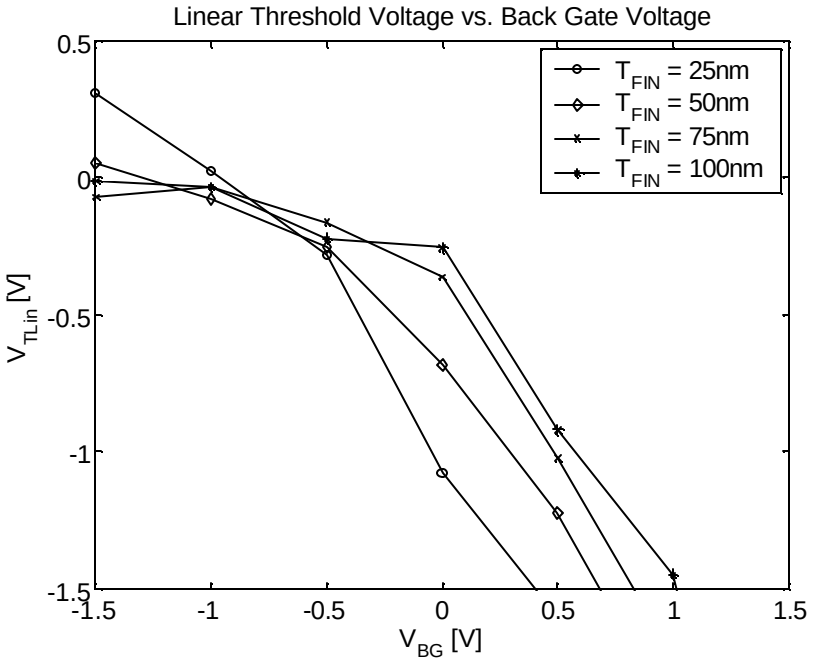

Figure 19. Simulated Linear Threshold Voltage vs. Back Gate Voltage for n-type IG-FinFETs of various fin thickness

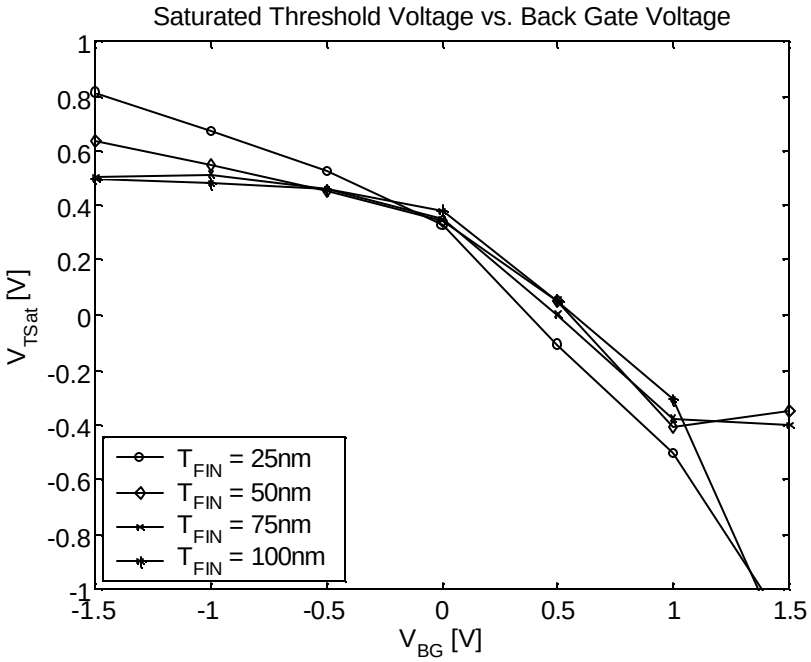

Figure 20. Simulated Saturated Threshold Voltage vs. Back Gate Voltage for n-type IG-FinFETs of various fin thickness

Essentially, what can be seen from the simulation results is that the devices with the thinnest fins exhibit the largest range of threshold voltage on the front channel. Since the threshold voltage is difficult to extract once the back channel is heavily inverted, the data gets somewhat clouded at the higher back gate biases (in the case of the NFET). However, in both cases of linear and saturated threshold voltages, the thinner fins show larger ranges in the independent gate device simulations as seen in both Figure 19 and Figure 20.

2.8 Effects of Source/Drain Junctions

One final new effect of double-gate devices that must be briefly discussed is the inherently three-dimensional nature of the source/drain junction profiles and the effect of these junctions on the performance of the composite device. In planar CMOS, the junction profiles serve to define, essentially in two dimensions, the barrier region between source/drain regions and the channel of the FET. The profile of the junction

below the surface of the FET affects the parasitic resistances and capacitances of the device and the effect of geometrical variations. In double gate CMOS, the implant profile has the additional effect of defining this barrier region for two different gates. In planar DG-CMOS, the natural graded junction profile defines the difference in metallurgical channel lengths and therefore overlaps between the top and bottom gates as seen in Figure 21. Since vertical implant profiles are nearly impossible to fabricate, this effect is nearly inescapable in planar DG-CMOS.

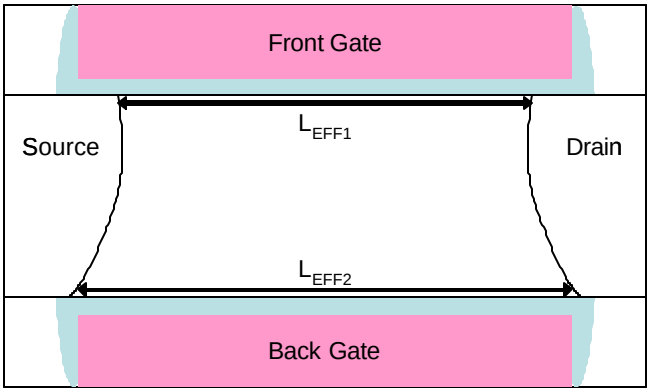

Figure 21. Effect of Source/Drain implant profile on effective gate lengths of double-gate device

In FinFET CMOS devices the problem is not solved, but changed. This natural implant profile can be somewhat controlled, as implants are typically done at high angles into the sides of the fin. Because of rotated quad implants, the effective gate lengths of the front and back gates can be well maintained. However, the top-to-bottom dopant distribution must also be carefully optimized. If more implant ends up in the top of the fin, and a deeper source/drain junction is formed at the top of the FinFET, then there will be a varying effective channel length from the top to the bottom of the fin, essentially across the active device width. This effect can manifest itself as a change in threshold from the top to the bottom of the fin, showing threshold voltage roll-off in one device. By having a lower threshold voltage at the top of the

fin, due to the shorter channel length, the device at the top of the fin will turn on first, while the bottom of the fin is still sub-threshold. This can be seen in measurements as a decrease in sub-threshold slope as the device will turn on slowly from top to bottom of the fin, instead of as a unified source-to-channel barrier. Maintaining this implant profile requires complex three dimensional implant modeling and tight process control. However, fabrication processes and integration schemes, such as angled implants and hardmask thicknesses, can be optimized to control this effect in FinFET CMOS.

2.9 Conclusion

Within this behavior lies the key to the independent double gate device. In certain conditions, it behaves as fully-depleted FET with active control of the threshold voltage. In other conditions, it operates as a linear current adder, with individual inputs. This versatile behavior may enable many different types of novel circuit design, ranging from analog circuits that would utilize the linear current mixing conditions for compact mixers, to adaptive digital logic that would use the back gate to change threshold voltages, changing the circuit from high-performance mode to low-power mode.

Various aspects of the physics that govern double-gate device behavior have shaped the fabrication processes utilized to realize these capabilities. In order to obtain a fully self-aligned double-gate FET that displays fully-depleted device behavior with independent control of the gates, the IG-FinFET has many advantages over other potential devices. Its single lithographic gate definition can eliminate gate-to-gate misalignment. The angled implants required to define the source and drain regions can help to minimize the effects of junction profiles on the short channel effects. Several fabrication concerns also appear to indicate the superiority of the IG-FinFET, including the fact that the gate dielectrics are grown simultaneously, and that

much of the device fabrication is conventional top-down processing, amenable to large scale manufacturing environments.

Chapter 3: Independent-Gate CMOS Circuit Design Implications

3.1 Overview

This chapter focuses on the applications and implications of using the Independent-Gate FinFET in circuit designs. First, a circuit design example is presented with circuit modeling results based on a quasi-static double-gate compact model. The results of these circuit simulations show the applications available to the IG-FinFET, and the potential leverage provided by threshold tunable CMOS. Nominal FinFET circuits are fabricated designed differently from conventional planar CMOS circuits, however, the circuit design and layout is nearly identical. With the introduction of the IG-FinFET, circuit design, and especially layout must be altered significantly. The second section of this chapter examines the differences between circuit designs and layouts of conventional planar single-gate, nominal FinFET, planar double-gate and IG-FinFET CMOS circuits. An analysis of essential design rules is presented with a scalable technology design parameter applied to the rules of interest. The chapter concludes with an assessment of the potential for application and integration for the IG-FinFET into VLSI CMOS designs.

3.2 Independent Gate Circuit Simulations

3.2.1 Device Compact Model

Device simulations as presented in Chapter 2 are excellent for evaluating individual devices and expectations for their DC characteristics. However, to evaluate the characteristics of even a simple two transistor inverter under many different operating conditions, a device simulation is quite cumbersome. A compact model for the double-gate transistor was required for any circuit simulation. A FORTRAN model was obtained from Dr. Paul M. Solomon at the IBM T. J. Watson Research Center [25]. This model essentially segments a single double-gate device into several standard FET models, connected them and biased them appropriately based on model device parameters and the bias conditions placed on the top-level DGFET. Based on applied bias conditions, the DGFET model segments the DC channel current into several conditions as seen in Figure 22: SINGLE, when either the front or back channel is providing the dominant current source, DUAL, when both channels are contributing a significant amount of current to the aggregate device current, and COMPOUND, an intermediate condition, where the channels are asymmetrically providing a significant amount of current, and the potential for charge sharing between the channel exists. This COMPOUND function serves well to model the effects of asymmetric double-gate devices, where because of asymmetric gate dielectrics, or gate workfunctions, the two inversion layers form under different conditions [25]. Because of this COMPOUND condition, this model can be effective at modeling the DC effects of biasing the two gates of an IG-FinFET asymmetrically.

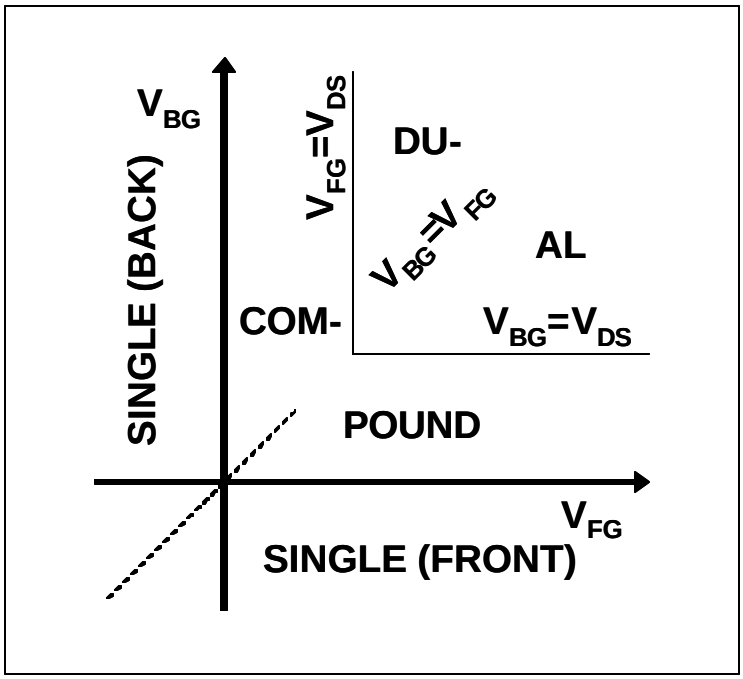

Figure 22. Partitions of the VFG/VBG plane according to the region of operation of the DGFET [25]

This model was converted and compiled to function in an available simulation environment. Several model parameters were parametrized to allow the circuit designer to define values for body thickness and oxide thickness. The oxide thickness, due to the IG-FinFET fabrication scheme used was applied symmetrically to both the front and back gate oxide thicknesses in the DGFET model. For the devices being fabricated this is appropriate given that both dielectrics are grown simultaneously. Other integration schemes allow the decoupling of this growth, which would require decoupling of these parameters for proper compact modeling. The extension lengths (Gamma) were also implemented with parametrized values, and different parameters associated with the source and drain extension lengths. Although all test structures developed in this work were designed for symmetric extension length, it is obvious that circuit designers would want to have the capability to design circuits with asymmetric extension lengths, for performance criteria, or layout issues. The gate

electrode workfunction was set at mid-gap (0.56eV) for both front and back gates on both NFETs and PFETs. This mid-gap setting simply affects the absolute threshold voltage of the devices, not the underlying double-gate nature of the devices. Setting this value to mid-gap is also consistent with many approaches for metal gate integration for T_{INV} scaling at low supply voltages [26].

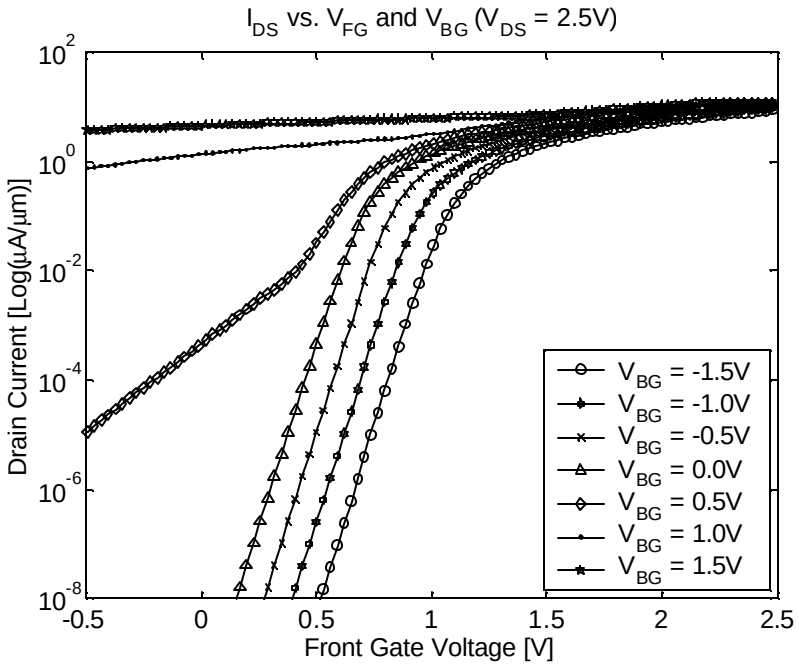

Figure 23. Simulated I_D vs. V_{FG}/V_{BG} curves for an NFET of T_{FIN} = 100nm, L_{GATE} = 2um, and symmetric source and drain extension lengths of 2um, using the converted, adjusted compact DGFET model.

Simple I_D vs. V_{FG} and V_{BG} curves, as measured on fabricated devices, produce curves as shown in Figure 23. Qualitatively, these curves match device simulation results quite well, with the obvious exception of the absolute threshold shift resulting from the mid-gap metal gate model setting.

This model is a quasi static model, but associates resistance to each terminal as a function of extension length, fin thickness and fin height (calculating a cross sectional area) multiplied by a bulk resistivity of silicon based on the source drain

doping concentration. Also, capacitances are calculated based on the charges on each node of the internal compact model.

3.2.2 DC CMOS Inverter Circuit Simulation and Results

A simple Independent Double-Gate CMOS inverter is designed according to the schematic shown in Figure 24, using the front gates of both devices as the "input"node and the back gates of each device separately as threshold adjustment nodes.

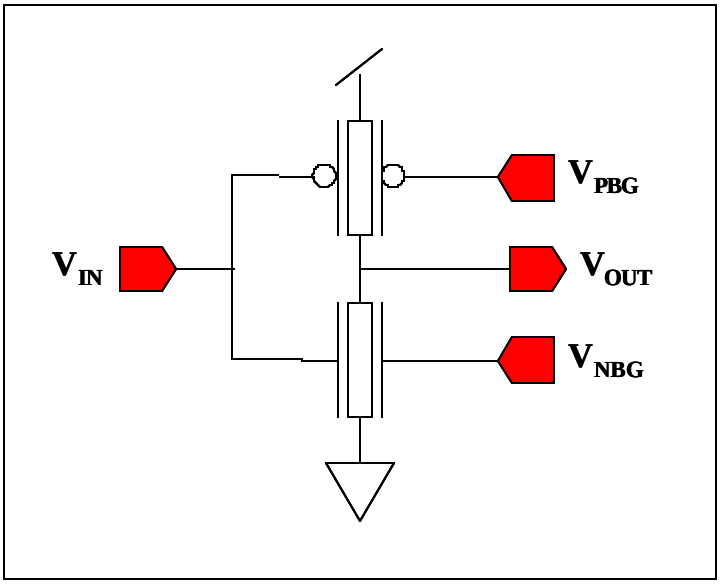

Figure 24.Circuit schematic of Independent gate CMOS inverter

This CMOS inverter was simulated in a wide range of bias conditions and the simulation results were compiled to show contours across the range of V_{NBG} (NFET Back Gate Voltage) and V_{PBG} (PFET Back Gate Voltage). V_{DD} was set to 2.5V and the back gate biases were modulated from V_S - V_{DD} to V_S + V_{DD} for each device. Therefore the range for V_{NBG} is -2.5V to 2.5V and the range of V_{PBG} is 0.0V to 5.0V. The first observation of the simulation results is that under certain back gate bias conditions, the output voltage will not swing from rail to rail as seen in Figure 25.

This is the case where the device that would typically be "off" in a single-gate circuit is biased by the back gate into the regime where the front gate can not turn the aggregate device current off. This range is mainly determined by the absolute threshold voltage setting in the model parameters. For a CMOS logic application, the region of interest for the operation of this circuit is the range where the output is allowed to swing from rail to rail as indicated by the shaded region in Figure 25.

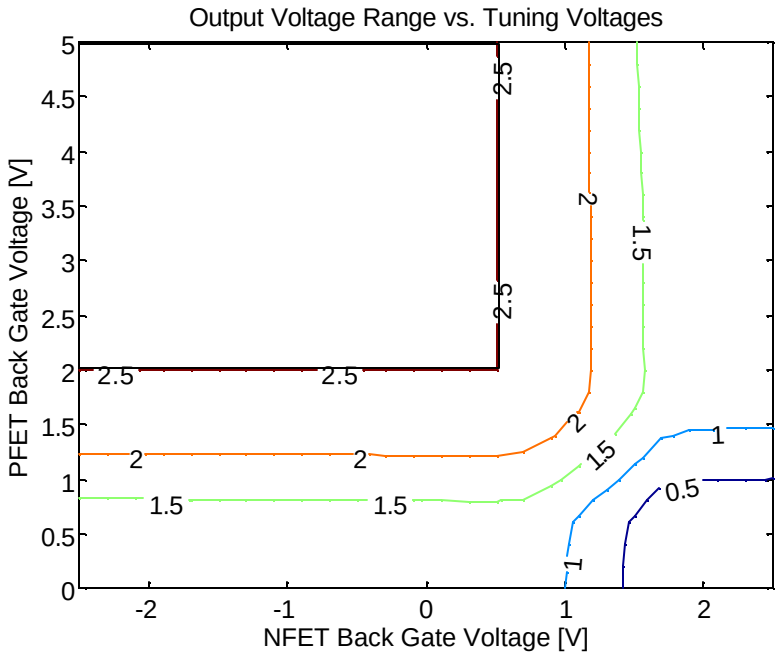

Figure 25. Output voltage range as a function of back gate voltages for IG-CMOS inverter. The shaded area shows the regime where output voltage swings from rail to rail.

As expected the peak switching current increases in this area of interest for CMOS design in the direction approaching the areas where the circuit is unable to swing from rail to rail as seen in Figure 26. Past the point where the circuit does not swing across the full range, the peak switching current increases dramatically. In this region however, the circuit is not acting like a CMOS inverter, but more like a resistively loaded inverter, and the increase in current is expected. Within the region

of interest, the peak switching current transitions through one full order of magnitude in current, showing a strong dependence on the back gate voltages. The potential application of this dependence is the leverage of back gate bias to change a CMOS circuit from a high performance mode (driving large current to the following stage) to a low power mode (driving small current and sacrificing performance for active power consumption). This concept is scalable to the idea of a large circuit or even an entire chip transitioning between these modes based on external biases.

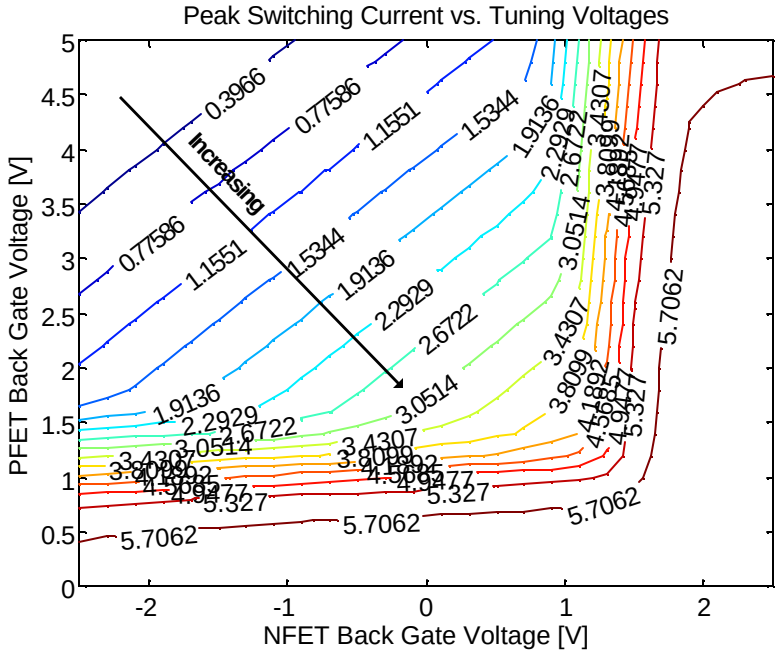

Figure 26. Peak Switching current as a function of back gate voltages. The arrow indicates the direction of increasing peak switching current in the region of CMOS circuit interest.

As seen in Figure 27, the transition or switch voltage of the IG-CMOS inverter also has a strong dependence on the back gate voltages. In planar CMOS, this value is determined by the threshold voltages of NFET and PFET and the "beta ratio", the ratio of the NFET device width to the PFET device width. Once the devices are chosen, and the circuit is laid out, there is no variation of the switch voltage in planar CMOS.

This IG-CMOS simulation was set up to simulate the devices as single fins only, so the device widths are equivalent. Yet, the addition of the independent back gates still allows for the variation of the switch voltage. The direction of increase in the switch voltage trend is 90º offset from the trend direction in the peak current trend. This means that by independently modulating V_{NBG} and V_{PBG}, these two important metrics of the IG-CMOS inverter can be varied somewhat decoupled from one another.

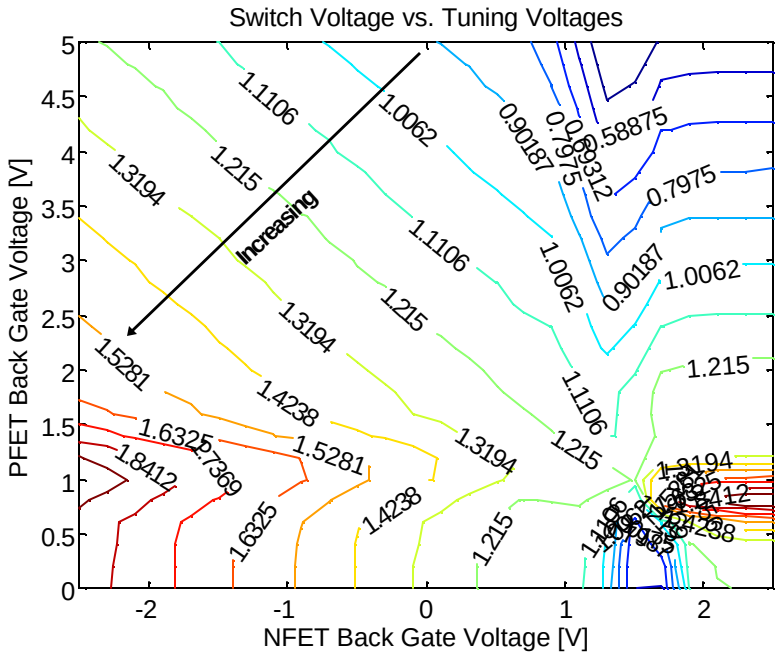

Figure 27. Inverter transition voltage as a function of back gate voltages. The arrow indicates the direction of increasing switch voltage in the region of CMOS circuit interest.

3.2.3 CMOS Ring Oscillator Circuit Simulation and Results

By connecting several inverters together in a ring, connecting all of the NFET back gates together and connecting all the PFET back gates together, a tunable ring oscillator can be designed. This is illustrated in Figure 28. A 19 stage IG-CMOS ring oscillator was simulated using the previously described compact model under various bias conditions.

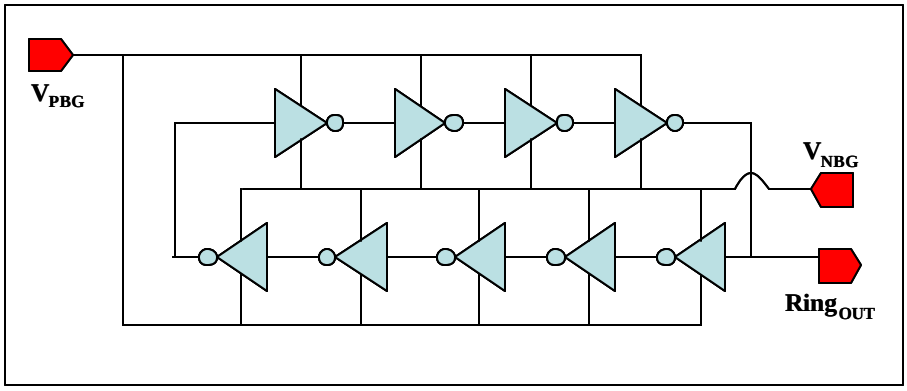

Figure 28. Schematic representation of a IG-CMOS ring oscillator with all PFET back gates all tied to V_{PBG} and all NFET back gates tied to V_{NBG}

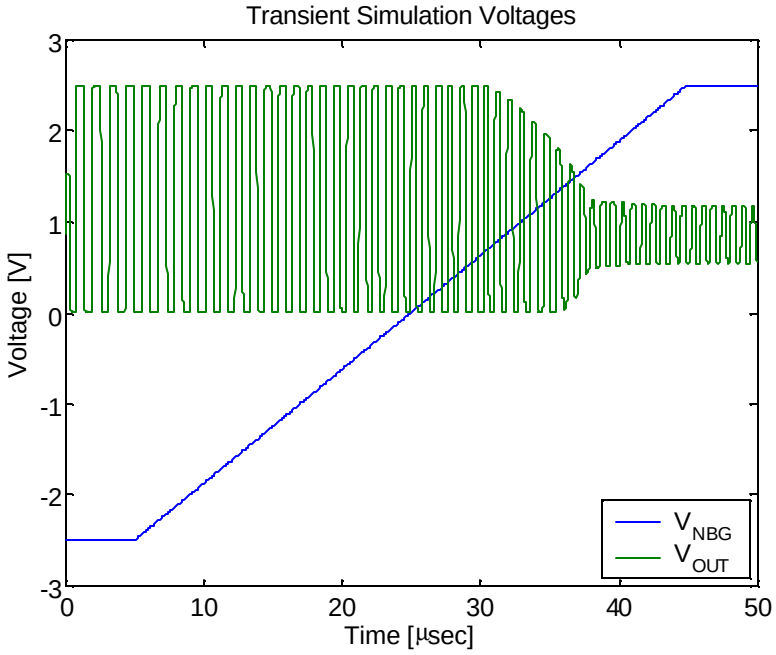

Figure 29. Results of transient simulation of 19 stage IG-CMOS ring oscillator under varying NFET back gate voltage

The results of the transient simulation, as shown in Figure 29, illustrate some of the interesting applications of this type of circuit. First of all, in the first half of the simulation time, the oscillation frequency is increasing as V_{NBG} is increased. This is expected due to the increased current flowing in the NFET while in saturation state. The period of the ring is extracted, and plotted in Figure 30. As V_{NBG} approaches

1.0V, the ring oscillator output begins to distort, not reaching the supply voltage. This is due to the fact that, with this high bias on the NFET back gate, the NFET never turns completely off, essentially resistively loading the pull-down device. A significant amount of current is still flowing through the NFET, pulling the "high" state at the output node of the ring down below the supply voltage. Finally, when the output voltage swing falls below 1.5V on the high side, the PFET is not properly turning off, due to decreased output voltage swing. At this point, the output voltage swing fails to reach ground, due to the increased PFET current flowing into the output node. The back gate voltage values when these types of swing failures occur are determined by the absolute threshold voltage, as set in the model.

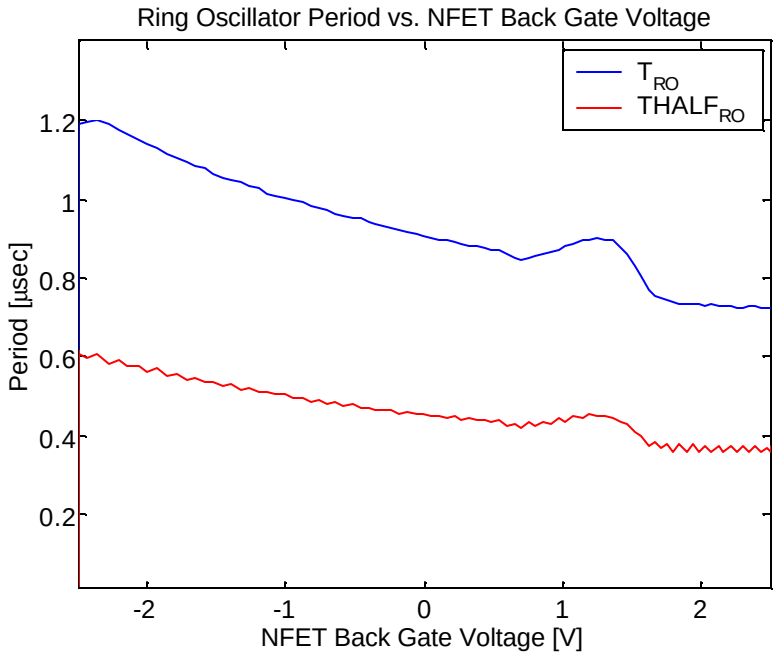

Figure 30. Extracted simulated ring oscillator period vs. V_{NBG} for 19 stage IG-CMOS ring oscillator

The extracted period shows interesting behavior. As expected and observed in the transient simulation results, the period of the ring is decreasing with increasing

NFET back gate voltage. However, there is an abnormal shape in the data near where the swing failures occur. As the output voltage swing begins to decrease due to the NFET not turning off, the period actually increases slightly because the voltage between the drain and source of the NFET has dropped, decreasing the current in the NFET during its on-state. Once the PFET begins failing to turn off, the output voltage swing is reduced again, and the period returns to decreasing behavior, as the total swing has become very small. A complementary analysis indicates the same type of circuit behavior when the PFET back gate is adjusted.

3.3 Design and Layout Considerations

These circuit simulations show several interesting design applications that can be built into VLSI CMOS. If one gate can be used to control the current drive of a static CMOS element, large scale designs could be biased into different operating modes. A high-performance mode can be achieved by increasing the current drive of the circuit. This would come at the expense of both static and dynamic power dissipation, as both saturation current and leakage are increased with back gate bias. A low-power mode can be achieved by decreasing the current drive of the circuit, at the expense of performance. This would allow the performance and short channel control of thin silicon CMOS with the circuit control of body contacted bulk CMOS. All of these circuit design modifications come with some expense of layout area due to the additional gate contacts required.

3.3.1 Layout Area Scaling

Layout area is at a premium in VLSI circuit designs. In conventional device scaling, the layout dimensions are scaled in unison with the channel length. The increase in current density of scaled devices allows for equivalent currents in smaller device widths. These two effects lead to much smaller circuits. When advanced

device structures are introduced, they are typically accompanied by new design rules. These design rules, and their impact on CMOS circuit layouts, must be examined prior to large scale design integration. To assess the impact of FinFET and IG-FinFET layout rules, an example set of parametrized planar CMOS design rules is suggested in the first portion of Table 2. A single device layout generated using these ? = 100nm rules, as shown in FIG, utilizes $0.7um^2$ (1.4um x 0.5um) to achieve 1um of design device width.

Table 2. Example of parametrized design rules for planar CMOS layout. Values are also given for ? = 100nm.

Rule	*Parametrized*	*? = 100nm*
Planar CMOS		
PC Width (Gate Length)	?	100
PC Corner to RX (same FET)	?	100
CA size	?	100
CA to CA space	1.5*?	150
CA to PC (same FET)	0.5*?	50
CA within RX (2 sides)	0.5*?	50
CA within PC (2 sides)	0.5*?	50
FinFET		
RX to RX (Fin to Fin)	?	100
RX width (Fin)	0	0
RX Corner to PC (same FET)	?	100
IG-FinFET		
CA over PC to RX	?	100

With the introduction of the FinFET, a few more rules are required. First, the drawn fin dimension is assumed to be sub-lithographic, either written by e-beam as in this work, using sidewall image transfer [27], or fabricated by some other means of patterning. The critical dimension for the fins is pitch, and for this example, the minimum pitch is assumed to be ? = 100nm. The other important rule that affects FinFET layouts is the “RX Corner to PC” rule. This rule is driven by the ability to fabricate dense fine lines for a wide device area, and the desire to keep the resistive

extension length short. These rules are also summarized in Table 2. For this example, the "RX Corner to PC" rule is assumed to be the same as "PC Corner to RX", and therefore ?/2 = 50nm. A typical wide FinFET layout designed with these rules is show in Figure 31b.

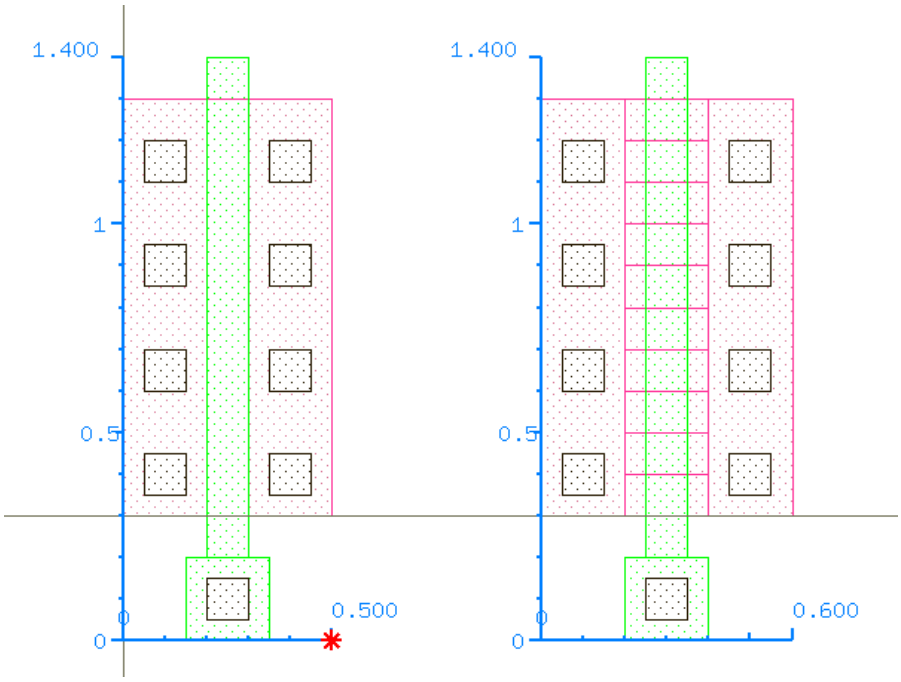

Figure 31. Planar CMOS layout (a - left) and equivalent FinFET layout (b - right)

This device layout is identical in the width dimension to the planar CMOS layout, but due to the introduction of the "RX Corner to PC" rule, is larger in the length dimension by ? = 100nm. This FinFET layout uses 0.84um^2 (1.4um x 0.6um) to achieve 11 fins. Assuming that the effective width of a FinFET is twice the height of the fin, the effective width of this layout is 22 x H_{FIN}. In this example, as long as the H_{FIN} is equal to 1um / 22 = 45nm, then the effective width per layout width is the same. The fin height is mainly limited by fabrication processes, and specifically the shadowing of angled ion implants. An array of fins at 100nm pitch with 45nm height is below the regime of this shadowing. The fin height could therefore be increased to

deliver a larger effective device width per unit layout width, if required. In order to achieve the same effective device width per layout area, the fin height would only have to be increased to 55nm. Any further increase in the height of the fin would result in more effective device width per unit layout area than the planar CMOS layout as shown in this example. In this manner, a FinFET layout can actually be more area efficient than a planar CMOS layout.

With the introduction of the IG-FinFET, comes another problem and more associated design rules. The problem arises from the fact that, after the separation of the gate electrodes, each gate segment must be contacted electrically. Therefore, a CA must be placed on the gate poly line between each pair of fins. A major design rule affecting the IG-FinFET layout is the "CA over PC to RX" rule. This rule forces the fins to be designed with spaces three times larger than the nominal FinFET layout, in order to place a contact on every gate segment between every pair of fins, as seen in Figure 32. With just this increase, the size of the single device changes to 1.232um^2 (0.7um x 1.76um) to provide just 5 fins. In addition, the effective device width in most IG-FinFET applications where one gate is used to adjust threshold and the other gate is used to modulate current, is not twice the height of the fin. For this example, the effective device width is still normalized to twice the fin height, but this is optimistic. As a result, the fin density per layout area has gone from 13.095 fins/um2 (11 fins / 0.84um2 in nominal FinFET layout) to 4.058 fins/um2 (5 fins / 1.232um2), more than a 3x reduction in area efficiency, without taking into account the reduced effective device width in IG applications.

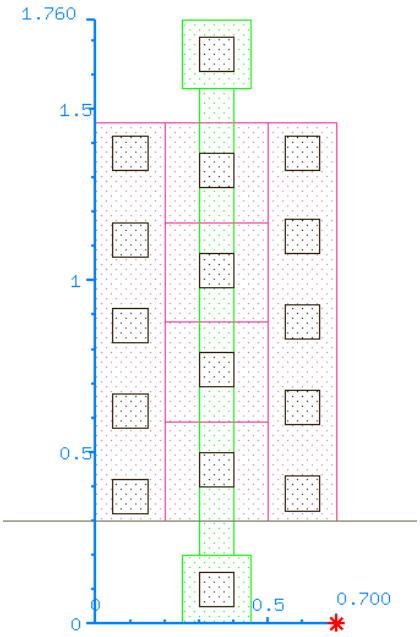

Figure 32. IG-FinFET single device layout

This area efficiency penalty is made even worse when the first level of metal is drawn. The device shown in Figure 32 is nearly impossible to wire properly as a true IG-FinFET, where every other gate segment should be connected together. Of course, interesting circuits can be designed using many inputs to the different gate nodes of this design, but these could also be achieved with several parallel single-fin IG-FinFET devices that do not share a source and drain. To add M1 comb-type structures to connect alternating gate segments would force the S/D contacts to be 3? from the gate contacts, currently only 1.5? away. This adds 1.5? to each side of the device, or 3? to the length dimension of the layout. This increases the layout area to 1.76um^2, which represents another 43% layout penalty. Another contact scheme would make small M1 contacts to the gates and connect them on M2, to preserve some of the layout density. However, this solution severely limits wiring flexibility and increases process complexity by potentially introducing additional BEOL levels.

3.4 Conclusion

Based on circuit simulations utilizing a quasi static compact model, it is quite apparent that IG-CMOS offers behavioral characteristics than can be leveraged for system applications. The ability to adjust the threshold of one or more devices in a circuit, while still maintaining the performance and short channel control of thin body CMOS may enable novel power management technology, or increased chip functionality. The IG-FinFET is one of very few devices that can be manufactured in a conventional CMOS manufacturing style, offer the IG CMOS benefits and deliver fully-self aligned gates. However, the layout penalty for using these devices is quite substantial. A 3x decrease in layout efficiency is clearly unacceptable for VLSI CMOS design. It is apparent that the application for this device is in a technology where nominal FinFETs could be integrated as well, offering increased layout efficiency, and the behavioral characteristics of IG-FinFETs where required.

Chapter 4: Independent-Gate FinFET Testsite Design

4.1 Overview

This chapter describes the design of a process development testsite for the fabrication of Independent-Gate FinFETs. The goal of this process development is the integration and fabrication of IG-FinFETs with channel lengths ranging from 250 nm to 5 microns and designed fin thicknesses ranging from 10 nm to 100 nm, with complete gate-to-gate isolation. The complete gate-to-gate isolation is required in order to effectively utilize the IG-FinFET in independent-gate mode, with the two gates biased at different voltages. This chapter begins with a brief description intended process integration scheme, followed by an overview of the testsite design. The integration scheme described is general in nature, and not specific to given unit processes used. The intent of this subsection is to introduce the general structures and features necessary for overall integration, in order to explain given design decisions. Each design subsection details the design of specific structures, the critical dimensions and the intended purpose of the fabricated structures. Both electrical test features and process monitor features are discussed. More detailed images of the testsite design are shown in Appendix A.

4.2 Integration Scheme

The initial fabrication step in most FinFET integration schemes is the definition of the silicon Fin body. Utilizing an SOI wafer allows the fin height to be

determined by the original SOI thickness. While embodiments of FinFETs utilizing Bulk Silicon wafers exist [16], the SOI integration scheme is more common. Typically, an oxide hardmask is used to define and etch the fin. Wet cleans and a sacrificial oxidation and strip are used to remove Reactive Ion Etch (RIE) polymer and smooth the RIE damaged sidewalls of the fin. The gate oxide is grown and polysilicon is deposited immediately as the gate electrode. At this point, the processing for IG-FinFETs deviates from that of nominal FinFETs. The gates must be isolated at this point to allow for a single lithography step to define both gates. Since the goal of IG-FinFET integration is self-aligned gates, the gate lithography must be accomplished in one step. A Chemical Mechanical Polish (CMP) process is used to recess the gates down to the height of the top of the remaining oxide hardmask on top of the fin. A brief Silicon RIE ensures complete gate-to-gate isolation over the fin. At this point, a gate electrode hardmask is deposited. In nominal FinFET processing, another oxide film is used. However, this is not possible for IG-FinFET integration. A nominal FinFET relies on the gate polysilicon on top of the fin to block the self-aligned source/drain implants. Since this part of the gate electrode has been polished away in the IG-FinFET, another film must be left in its place. This integration scheme utilizes the gate electrode hardmask to block the source/drain implants. Since the source and drain regions must be exposed during the source/drain implants, an oxide etch will be required after the gate etch. Therefore, the gate electrode mask, which must remain in place after the source and drain are exposed, must not be oxide. This film must be composed of a material that will resist both the polysilicon RIE to define the gate and the oxide etch to open the source and drain. This integration scheme utilizes Silicon Nitride (Si_3N_4) for the gate electrode hardmask. Once the gate electrode is defined and the source and drain areas are exposed, a sidewall reoxidation is used for two reasons. First, this helps to eliminate gate shorts that can form during

the gate etch. Second, this thin oxide will prevent the fragile fin extensions from being sputtered by the ion implant beam. Self-aligned source/drain implants follow. This process flow, including the deviations required to form the Independent-Gate FinFET, is shown graphically in Figure 33.

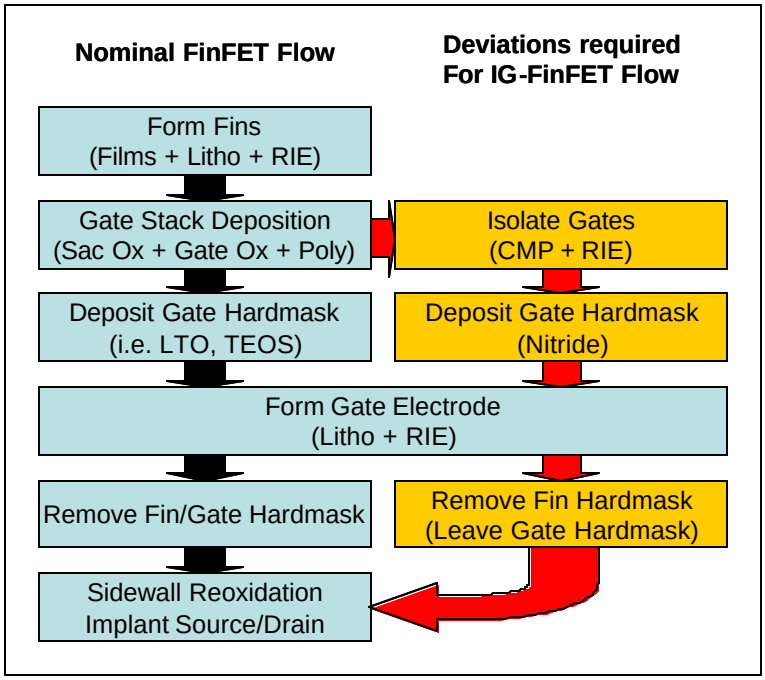

Figure 33. Process flow for nominal FinFET fabrication and deviations required for Independent-Gate FinFET fabrication

The final process module is a simple metallization process that uses a contact etch, Aluminum sputter deposition, and first-level metal etch. A final oxide passivation layer protects the metallization, and probe openings are etched. A final low temperature anneal is used to eliminate surface states in the FETs and to improve Aluminum to Silicon contact resistance.

4.3 Testsite Design

The main objective of the testsite design was to provide fabrication process monitors and testable access to process monitors, individual devices across a large

range of device dimensions, and simple circuits that demonstrate novel applications of independent-gate behavior.

4.3.1 Single Devices and Arrays

The first two critical dimensions to single device design are the fin thickness (T_{FIN}) and gate length (L_{GATE}). Both of these dimensions are defined lithographically. One other dimension of interest is the distance from the edge of the gate to the point of the fin that flares out for a source/drain contact. Between these points, the fin extension (L_{EXT}), and therefore the source or drain, is extremely thin and highly resistive. This resistance is external to the intrinsic device, and adds to the overall series resistance of the extrinsic device. These three critical dimensions are shown in Figure 34.

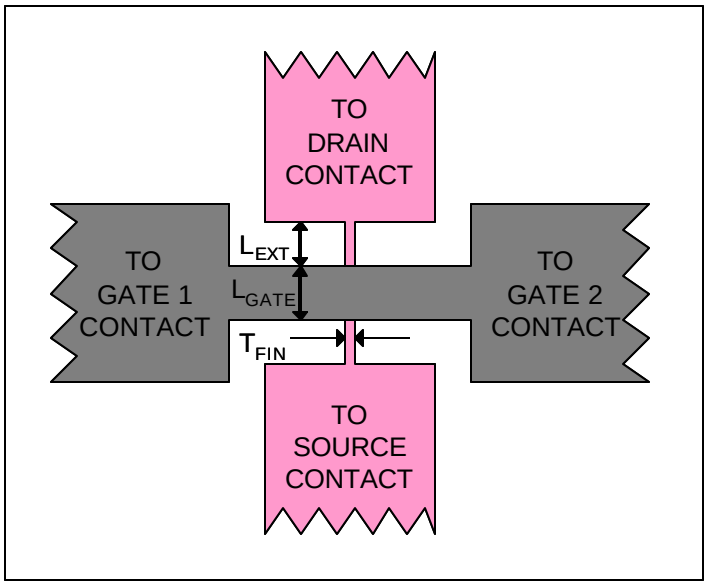

Figure 34. Three critical dimensions in IG-FinFET individual device layout

Making L_{EXT} small requires precise overlay alignment of the gate mask to the existing fin shape. Easing this overlay tolerance, and expanding this dimension, increases the series resistance. Figure 35a-b shows the effect of a minor overlay error

on devices with long L_{EXT} and short L_{EXT}. The device with the longer L_{EXT} will still be a functional device after the overlay error, but will have higher extrinsic resistance than a device with shorter L_{EXT}. The device designed to have shorter L_{EXT} may not even be functional after the overlay error. Therefore this dimension controls a trade-off between yield and performance. An optimal technology would offer a self-aligned method of integrating the source/drain flare as shown in Figure 35c. In fact, a great deal of work is being done in industry for self-aligned raised source drain for planar CMOS. This technology can also be applied to FinFET CMOS. However, this aspect of the integration is beyond the scope of this work, and the capabilities of the CNF.

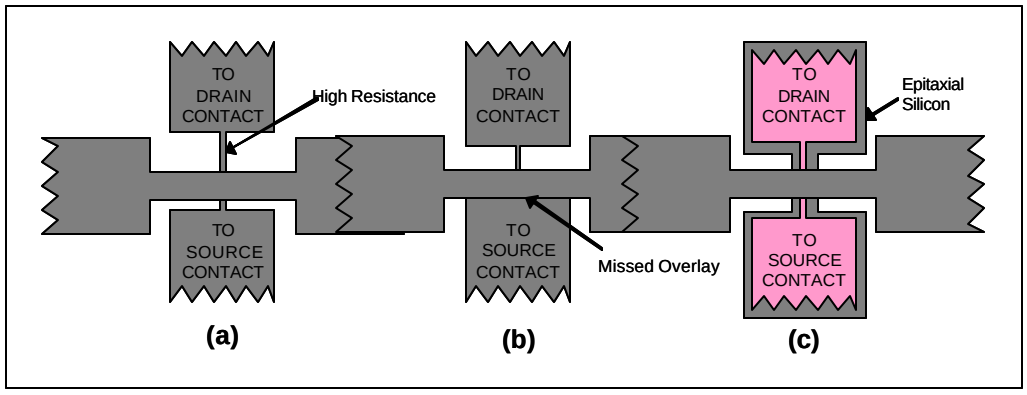

Figure 35. Effect of extension length variation for overlay tolerance

Based on these three critical dimensions, devices were designed across a large range of parameter space. The fin thickness is the smallest critical dimension. Because of this, electron beam lithography was chosen as the most promising method of definition for the RX (Active) level. The CNF LeicaVB6 has been shown to have patterning capabilities near 5nm resolution. To provide a large process window for variation, devices with fin thicknesses of 10, 25, 50, 75, 100 nm were designed. In order to provide adequate alignment tolerances, the Nikon NSR DUV lithography system was chosen for the PC (gate electrode) lithography level. The alignment is

done automatically by this tool, and it has been shown to have resolution capabilities near 250 nm and overlay tolerances near 100 nm. Because of these specifications and in order to measure short and long channel electrical characteristics of the IG-FinFET, devices with gate lengths of 0.25, 0.35, 0.5, 2.0, 5.0 microns were designed. Extension lengths of the same dimensions were chosen (0.25, 0.35, 0.5, 2.0, 5.0 microns).

Outside the intrinsic device, the gate (PC), source and drain (RX) areas for probing and contacts were made extremely large (40 micron x 40 micron). This was meant to facilitate probing the device after source drain implants without metallization. Additional levels were designed to provide metal probing contacts for improved contact resistance. The metal pads were designed in octagonal shapes on M1 (metal 1) level, to attempt to minimize shorting, with 40 micron x 40 micron dimension. The contacts to tie the metal pads to the gate, source and drain areas were designed at 5 micron x 5 micron dimension and placed in 2 x 2 arrays in CA (contact to active) level on each pad. The testable layout for a single individual device is shown in Figure 36.

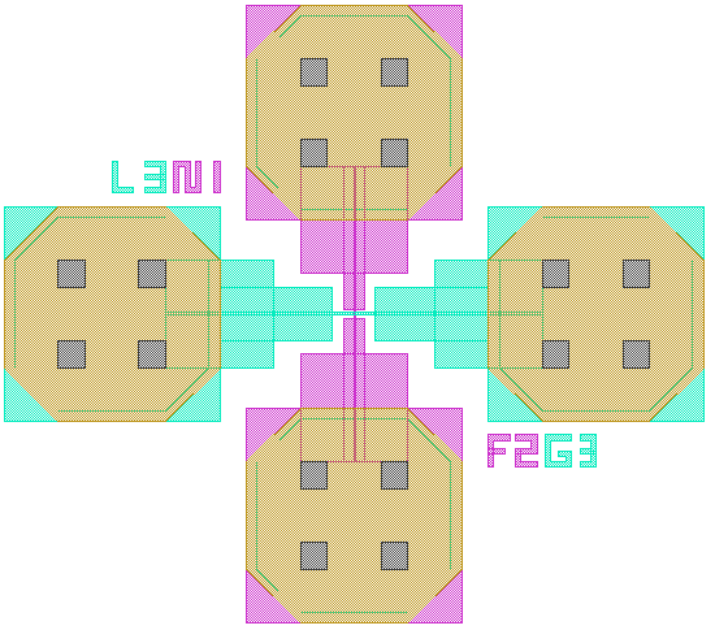

Figure 36. Testsite layout of single IG-FinFET (Source/Drain are top/bottom, Front/Back Gates are left/right). Octogonal metal pads are designed to optimize packing of devices into an array.

In order to make the design area small (to minimize across-chip lithography variations), the individual devices were arrayed in the following manner. Arrays of 25 devices were arranged to vary with gate length in the x-direction and extension length in the y-direction. Adjacent devices share one contact. This is shown in Figure 37. This array was copied five times onto fins of the different thicknesses. The entire macro array was duplicated to provide an entire set of testable NFETs and PFETs for CMOS integration. This was accomplished with a single block mask for each device type. Since the integration scheme is to have self aligned source and drain regions, the block level lithography is not extremely critical. The NFETs are surrounded by BN level, the PFETs by BP level. This set of dimensions results in 125 NFETs and 125 PFETs (5 fin thicknesses, 5 gate lengths, 5 extension lengths).

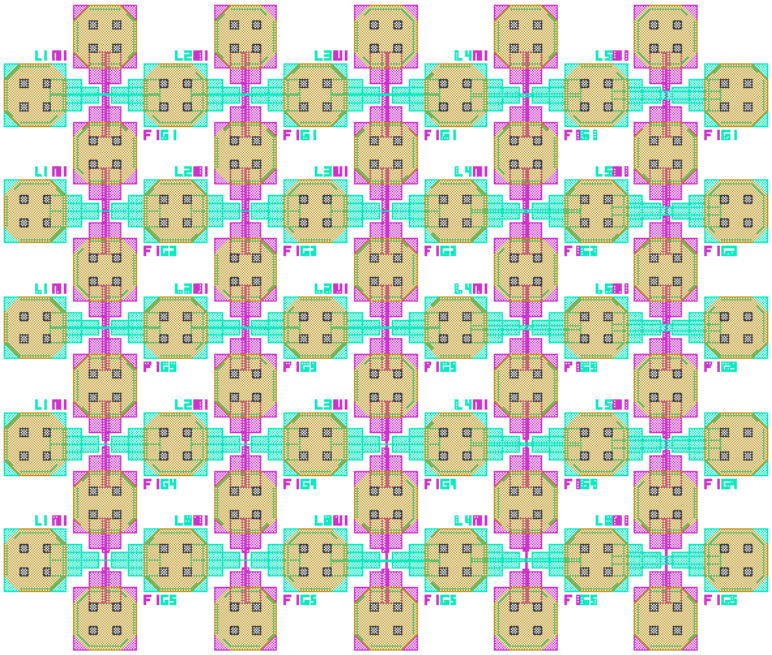

Figure 37. Testsite array of 25 individually testable single IG-FinFETs

4.3.2 Independent-Gate Inverters

The independent gate inverter is the simplest of all logic gates that can demonstrate the unique behavioral characteristics of the IG-FinFET. The circuit is designed identically to a nominal inverter, with the NFET and PFET sharing a source/drain node. In the IG-FinFET version of the inverter, the back-gates of both NFET and PFET are tied out to separate contacts. Having these separate contacts allows the circuit to be adjusted by two bias points. These bias points can modulate the switching point, the current and noise margins of the inverter. Simulation results for this type of circuit are discussed in Chapter 3. Minor layout modifications between the individual devices and the inverter devices were necessary. For circuit density and source drain resistance concerns, the large source/drain contact regions were designed significantly smaller (10 microns x 10 microns). These regions are large enough to land only one 5 micron x 5 micron contact. The gate contact regions were reduced in size by the same amount. Due to the significant reduction in the amount of active

silicon region in the circuit, fill shapes were added around the circuit. These fill shapes are designed to assist in the planarization process that will be discussed later. Since the RX areas serve as the polish-stop, the reduction of RX area will make stopping the polish difficult. Over-polishing the poly could lead to dishing or the recessing of the gate poly below the height of the silicon fin. Finally, since no Silicide process was planned for these circuits, M1 was designed to bridge and short all necessary p-n junctions in the layout. This occurred at the front gate electrode which is a single poly shape shared by both PFET and NFET, and at the shared source/drain output node. The single inverter layout is shown in Figure 38. Inverters were designed across a slightly reduced set of critical dimensions. The 5 micron channel lengths and 5 micron extension length devices were eliminated. All inverters were designed with symmetric devices, meaning the NFET was identical in dimension to the PFET. This set of dimensions results in 80 individual inverters (5 fin thicknesses, 4 gate lengths, 4 extension lengths).

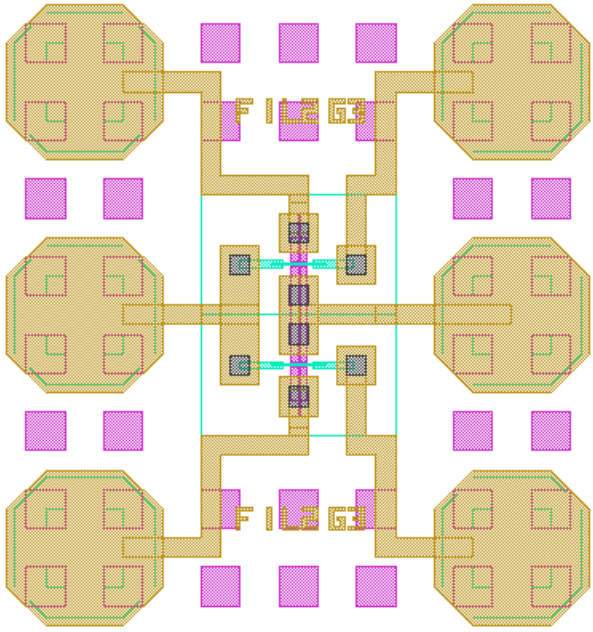

Figure 38. Testsite layout of IG-FinFET CMOS Inverter

4.3.3 Tunable Ring Oscillators

19 stage ring oscillators were designed using the identical inverter layout described in the previous section. All of the NFET back gates are tied together and all of the PFET back gates are tied together to provide for two bias point adjustments. The enable of the ring oscillator was designed by connecting the V_{DD} point of one of the inverters in the ring to a separate node. If the ring V_{DD} and ground are applied with the Enable node held to ground, the ring should hold a stable state. Once the Enable node it brought to V_{DD}, the ring should begin to oscillate. This ring oscillator design style makes additional circuit design unnecessary and preserves identical stages throughout the design. Each stage of the oscillator is "brick-walled", meaning laid out as close to one another as possible, to minimize inter-stage wiring capacitance. The layout for one full ring oscillator is shown in Figure 39. Because of chip size constraints, ring oscillators were designed across a smaller design dimension window. Fin thicknesses of 25, 50, 75 and 100 nm, Gate Lengths of 0.25 and 2 microns, and Extension Lengths of 0.25 and 2 microns were used. This resulted in a total of 16 ring oscillators (4 fin thicknesses, 2 gate lengths, 2 extension lengths).

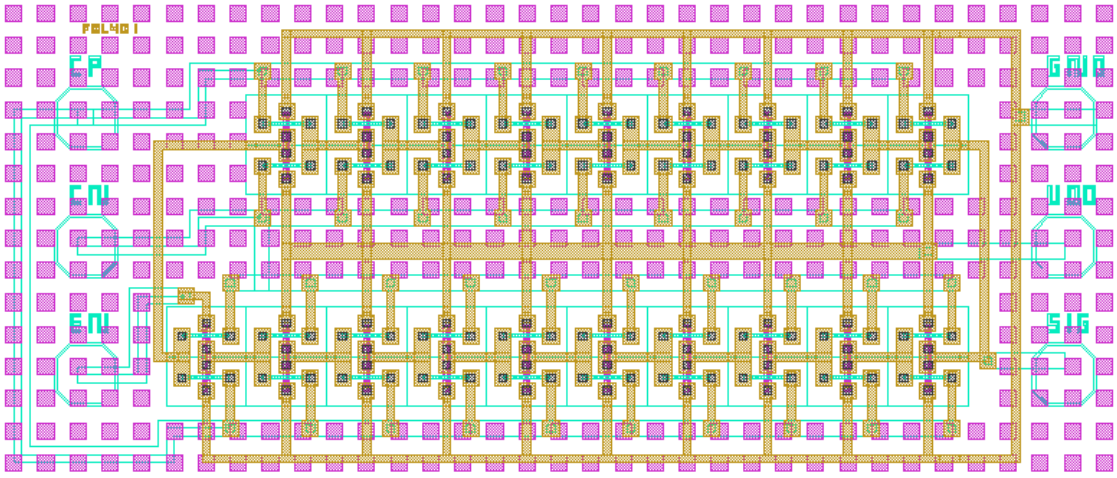

Figure 39. Testsite layout of IG-FinFET CMOS Tunable Ring Oscillator

4.3.4 Test Structures

In addition to FETs and circuits, the testsite was designed to include other electrical test and process measurement structures. Small arrays of fins were designed at several points around the chip to provide for areas to perform profilometry measurements. These profilometry measurements can be used to determine the step height of the exposed resist pattern, the etched oxide hardmask and the final etched fin. Due to the measurement method and the critical dimensions, this data would provide almost no data as to the sidewall angle of the resist, hardmask or fin. These measurements would have to be taken via cross-section SEM. A macro containing long lines of RX was designed to run the length of one edge of the testsite. These lines would provide an excellent location to cleave a sample to observe a cross section of the fin at any time in the fabrication process. 20 sets of 5 lines were designed, each set of five containing one line at each of the dimension in the FET fin thickness design window. These lines were anchored with large squares at each end of the macro, and one in the middle to provide mechanical stability to the lines. A PC shape was designed to cover half of the length of these lines. This half of the macro would provide a position to cleave a sample to observe a cross section of the gate polysilicon as deposited or polished during the process. The exposed area of the macro would provide a position to measure and observe the condition of the fin silicon external to the gate, where protecting the silicon during gate stack fabrication is critical to FET yield. Large rectangles in RX were designed at several points across the testsite to provide for a variety of measurements during the fabrication process. These features can be used for profilometry to determine RX step heights for large features. The features also provide adequate area to perform reflectance film thickness measurements. By using the Filmetrix optical reflectance film thickness measurement software, thicknesses of multiple films in a stack can be determined.

This is extremely useful for determining the remaining oxide hardmask thickness during several fabrication operations.

A critical operation in the fabrication of IG-FinFETs is the Chemical Mechanical Polishing of the polysilicon to isolate the two gates. The large features designed around the testsite can be measured via reflectance measurements to determine how much polysilicon remains on top of the oxide hard mask. However, this only gives a measure of the CMP progress over a large feature, while the CMP progress over the fin is the important data. A very simple structure was designed to provide a quick electrical measurement of polysilicon isolation from CMP. Squares of RX level were designed at thicknesses corresponding to the FET fin thickness design window. After polysilicon deposition, the polysilicon would be continuous over these squares from the inside to the outside. A simple resistance measurement between a micro-probe on the inside of the square and a micro-probe outside the square would indicate this continuity. As the CMP steps polished the top of the polysilicon away, this resistance would increase, until finally, upon completion of CMP, there would be complete isolation from inside to outside of the square. RX shapes were designed to provide Fin-type resistors. The resistor shapes were designed with identical contact shapes as the sources and drains of the FETs, across an identical window of fin thicknesses. Four sets of resistors were designed with different lengths, to allow for parametric extraction of resistance per unit length of the source/drain extension region. PC shapes were also designed to provide for polysilicon resistors. These resistors were also designed with identical contact schemes to the FETs, with widths varying across the FET gate length design window, to allow for parametric extraction of gate resistance.

Chapter 5: Independent-Gate FinFET Process Development

5.1 Overview

This chapter describes the development of a fabrication process for Independent-Gate FinFETs. Again, the goal of this process development is the integration and fabrication of IG-FinFETs with channel lengths ranging from 250 nm to 5 microns and designed fin thicknesses ranging from 10 nm to 100 nm, with complete gate-to-gate isolation. This chapter details each of the major process modules developed in order to integrate the IG-FinFET. Each subsection discusses the development challenges, process simulation results and physical analysis results including SEM images when appropriate. This process development was focused primarily on the development of a process for integration in the Cornell Nanofabrication Facility (CNF) with a secondary focus on the potential to transfer the fabrication process to other facilities. Because of this, several experiments and integration approaches were attempted to maximize the capabilities of the CNF, and minimize yield limiting factors. A more detailed process flow listing, including all processes required for CMOS integration, is provided in Appendix C.

5.2 Fin Definition and Etch

5.2.1 Challenges

The first module in the integration of FinFETs, both nominal and Independent-Gate, is the definition and etch of the fin body. Many challenges exist for this module, even though it is the most similar to a conventional process of any of the FinFET modules. The fin definition and etch is quite similar to a conventional planar CMOS gate definition and etch, in that extremely small cross sectional width is desired, with little to no line edge roughness. State of the art gate definition modules target the fabrication of gates below 90 nm thick. Ideally, the fin would be significantly thinner, on the order of 20 nm. FinFET integration also requires that the etch process be extremely anisotropic, to provide vertical sidewalls. Any angle in the sidewall profile leads immediately to a variation in the silicon thickness across the electrical width of the device. This, in turn, changes the threshold voltage across the width of device, deteriorating the subthreshold characteristics. IG-FinFET integration places another challenge into the development of the fin definition module. The polysilicon for the gates will be recessed below the height of the remaining hardmask oxide. This hardmask oxide must be thick enough at the time of polysilicon deposition to provide an adequate amount of oxide to isolate the gates, without the polysilicon being recessed below the height of the top of the silicon fin. This imposes requirements on the fin etch to be selective enough to leave an oxide hardmask remaining on the fin. This also places requirements on the clean operations between the fin etch and the polysilicon deposition to not remove this oxide hardmask.

Several integration schemes exist for fin definition and etch, each focused on different requirements. If optical lithography is to be used, as in a manufacturing environment, aggressive trimming must be used to reduce the thickness of the

hardmask or fin. If the fin thickness tolerance is to be addressed, schemes such as Sidewall Image Transfer (SIT) [27] can be used. This is an integration scheme whereby a polysilicon mandrel is defined, spacers are deposited and etched on the sides of the mandrel, and the mandrel is removed, leaving only the spacers, whose thickness are defined more by film deposition and etch than lithography, to mask the fin etch. Because of the electron beam lithography capabilities of the CNF, and the prototype nature of this work, direct pattern transfer using electron beam lithography was chosen for the definition of the fin.

5.2.2 Film Stack

Initially, the fin etch was envisioned to be the most difficult part of the integration scheme. To ease the difficulty in this area, a relatively short fin was desired. In the SOI embodiment of the FinFET, the fin height is determined by the SOI thickness under the hardmask. The etch difficulty can be characterized by the intended aspect ratio of the feature. The testsite design includes fins of thicknesses from 10nm to 100nm. If the fin were designed to be 100nm tall, this would result in an extremely difficult 10:1 aspect ratio for the 10nm designed fin, a difficult but not unreasonable 4:1 aspect ratio for the 25nm designed fin, and a relatively simple 1:1 aspect ratio for the 100nm fin. While the 10:1 aspect ratio on the 10nm fin may be out of the possible process window, the 100nm fin height makes for a wide array of aspect ratios in design. The hardmask for the fin etch needs to be quite thick for this process, as it will serve as the gate-to-gate isolation structure once the gates are polished. In order to give the CMP process some leeway in stopping the polysilicon polish in the correct position, a 100nm SiO2 hardmask is set as the initial target.

Commercially available 340nm 4” SOI wafers were purchased from SOITEC®. A standard MOS clean with a 10 second HF Dip was used to pre-clean the wafers and

remove any native oxide that may be present in shipping. A wet thermal oxidation was performed at 1000C for 66 minutes in the Thermco TCA Oxidation Tube #1 running recipe 50 (Steam Oxidation, No TCA, 7.1 l/min H2, 4 l/min O2) targeting the growth of a 450 nm film and thinning the SOI to 140 nm. This oxide was measured using the Filmetrix optical reflectance film measurement system. The Filmetrix has difficulty measuring stacks of many films, so monitor wafers of bulk silicon were run along with the SOI to measure the thickness. The results are shown in FIG. The monitor wafers, loaded in front and behind the SOI wafers, measured 335nm and 365nm of oxide respectively. The measurement also shows a distinct center to edge variation. This film was stripped in 49% HF for 2 minutes. Another wet thermal oxidation was performed at 1000C for 25 minutes with identical gas flows to target the growth of a 100 nm thick SiO2 hardmask, further thinning the SOI to 100nm. The wet thermal oxidation appears to have a large center to edge variation as large wafer to wafer non-uniformity. Because of this, the second oxidation was changed to recipe 20 (Dry Oxidation, 6 l/min O2, 0.24 l/min TCA in N2). This process required some development and the thickness targets were achieved on the third run. A summary of the film stack processing is show in Table 3. In the third run, a final SOI thickness of 98nm with an oxide hardmask thickness of 100nm was achieved.

Table 3. SOI Thinning and hardmask oxidation thickness results

Run	***1a***	***1b***	***1c***
Starting SOI Thickness	340	340	340
SOI Thinning Oxidation			
Recipe	Wet no TCA	Wet no TCA	Wet no TCA
Temperature [C]	1000	1000	1000
Time [min]	66	90	78
Measured Thickness [nm]	365	490	450
Silicon Consumed [nm]	160.6	215.6	198
Remaining SOI Thickness [nm]	179.4	124.4	142
Hardmask Oxidation			
Recipe	Wet no TCA	Dry w/ TCA	Dry w/ TCA
Time [min]	25	60	55
Temperature [°C]	1000	1050	1050
Measured Thickness [nm]	190	150	100
Silicon Consumed [nm]	83.6	66	44
Remaining SOI Thickness [nm]	95.8	58.4	98

5.2.3 Electron Beam Lithography

RX makes up a small percentage of the overall die area. Therefore, a negative resist will make the electron beam lithography less time consuming. NEB-31 is a negative electron-beam resist that may be diluted with PGMEA to provide multiple thicknesses of resist. A 5 minute 170ºC dehydration bake helped to prepare the surface for resist processing. To promote surface adhesion, P-20 primer was deposited on the wafer for 60 seconds, and then spun off at 3000 rpm for 60 seconds. For the 150 nm oxide hardmask, a thickness of 125 nm was required to provide adequate etch masking for the oxide RIE. This was accomplished using NEB-31 diluted 1:1 with PGMEA, spun at 4000 rpm for 60 seconds. Wafers were then baked at 110ºC for 2 minutes to remove the solvent. NEB-31 is quite sensitive to thermal budget, so strict adherence to the preparation technique was essential. Also, the resist spin and preparation was always performed within 30 minutes of loading the wafer into the Leica VB6.

The only critical area of the device that requires the resolution of electron beam lithography is the fin. However, the large source and drain contact regions need to be patterned in the same hardmask to allow for a unified device active area. A mixed e-beam and optical lithography scheme may be more efficient for this application, however, this would require an additional alignment operation that would increase the overall integration complexity. So, a single e-beam write was used to define the fins and the large source drain contact regions. The Leica VB6 has many different tuning options, but the main two that allow for very fine resolution are the Virtual Resolution Unit (VRU) and the exposure dose. The VRU determines the spot size of the electron beam, and the dose determines the speed that the beam passes across the surface of the wafer. In order to get the best possible resolution, the smallest possible spot size must be used. The smallest spot size selectable on the CNF Leica VB6 is 5nm using VRU=1. This would draw a 10nm line in two passes. The dose must be tightly controlled in the same manner as optical lithography, however, optical lithography uses an exposure time to control the dose, while e-beam lithography utilizes a raster speed control. Since the large source and drain contact regions are 40 microns across, writing them with a 5um spot would take an exceptional amount of time (8000 passes per contact region). In order to increase the efficiency of this approach, experiments were performed to fracture the RX level into two separate levels: one level for the fins, requiring the maximum resolution and one level for the contact regions with limited resolution requirements, optimized for tool time requirements. Once the design data was fractured into two files, a job file was written to expose the two files in one job, so that the two writes were aligned to one another. While the testsite was designed to provide many different fin thicknesses, the exposure can be modulated to increase the granularity of the experiment. Since each die is written independently, and the two fractured design data files are also written

independently, the dose can be varied from die to die to test the exposure and to vary the resulting fin thickness from die to die. Since the files use different VRUs, the doses are obviously much different. The tool is limited to a 25MHz raster frequency. For reproducible results, the raster frequency should be kept significantly below this limitation. The first write starts the beam at 1 nA and VRU 2 (10nm resolution) with a dose of 105uC/cm2 (a 9.5MHz raster frequency) and increments by 1uC/cm2 each die to a dose of 140uC/cm2 (a 7.1MHz raster frequency) on the last die. The second write starts the beam at 5nA and VRU 16 (80nm resolution) and a dose of 8uC/cm2 (a 9.8MHz raster frequency) and increments by 0.2uC/cm2 each die to a dose of 15uC/cm2 (a 5.2MHz raster frequency). The final job file to manipulate the beam dose as the files are stepped across the wafer is included in Appendix B. This job file writes a 6 x 6 array of the testsite in the center of the wafer. The die is roughly 4 mm x 4 mm and is written with a 5mm x 5 mm periodicity to leave a 1 mm dicing channel between die. The job file calibrates and sets up the electron beam then writes the fin RX data, sets the beam to a larger spot size and faster raster speed, refocuses, and finally writes the contact RX data, at incremented doses for both files.

Immediately after exposure, wafers were baked again to chemically amplify the exposure at 95ºC for 4 minutes. The wafer is cooled and developed in MF-321 for 1 sec for every 10 nm of resist. To ensure full development of 125 nm of resist, the development is done for 15 seconds. Proper exposure and development is easily monitored with an optical microscope observing the SEM macro as seen in Figure 40 or using the KLA Tencor P-20 Profilometer to measure the resist profile on large features.

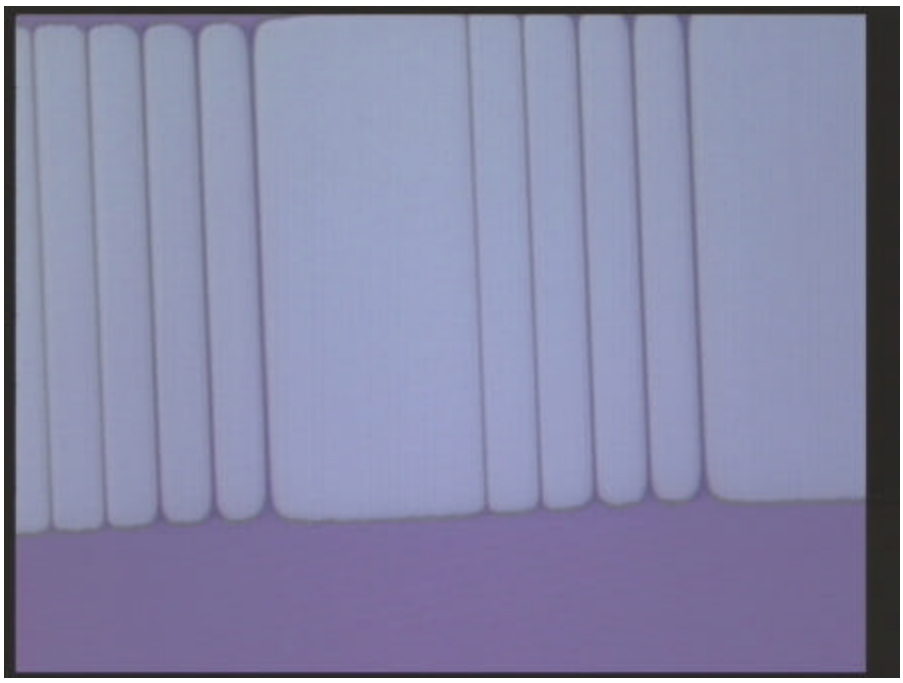

Figure 40. Top-down micrograph of SEM Macro, showing 10, 25, 50, 75 and 100nm line exposed in NEB-31

5.2.4 Hardmask Etch

Monitor wafers showed that the Applied Material RIE standard thermal oxide etch recipe at 30mT pressure, 30sccm CHF_3 at 90W of incident RF power resulted in a consistent etch rate of roughly 30nm/min with little center to edge variation while using the top gas feed. The PlasmaTherm PT-72 standard thermal oxide recipe showed less center to edge variation but the etch rate varied from experiment to experiment. Since all of the die are written in the center inch of a 4" wafer, the center to edge variation is less important than the consistency of the etch rate. The Applied Materials etch was run for 5.5 minutes, or an estimated 10% over-etch to ensure that the silicon was exposed between narrowly nested fin shapes. At this point, etch completion may be monitored by applying a droplet of water to the inactive silicon area on the wafer. If the oxide is cleared, the water will not adhere to the bare silicon. If the etch is not complete, and oxide remains, the water will adhere. The profile of

the etched fin hardmask is difficult to measure without a cross section. However, the height of the hardmask can be measured simply with the profilometer. The sparse SEM array is useful for this measurement, as the spaces between the fins are large, and the profilometer pin can reach between the fins. This profilometry measurement is used to determine the resist etch rate, and possible oxide erosion on both fins (in the SEM array) and larger features.

In order to reduce polymer contamination on the fin sidewalls, the resist must be removed prior to the silicon fin etch. While this will increase the erosion of the oxide hardmask (versus leaving the resist in place in a soft-mask process), the resist contamination on the fin sidewalls may be a more damaging effect. NEB-31 is relatively difficult resist to remove. The GaSonics Aura 1000 plasma etcher was used to run an O_2 plasma etch for 60 seconds with 20 seconds of heat (standard recipe #8) to sinter and etch the resist. A final clean in Acetone and Isopropyl Alcohol (IPA) removes all remaining resist and foreign particles.

5.2.5 Fin Etch

Since the sidewalls of the fin will form the Oxide-Semiconductor interface, the requirements on the etch process are stringent. An etch must be chosen to leave a minimum of RIE damage and etch polymers on the sidewalls. For this reason, the cyclic Bosch® etch was not acceptable. Because the fin has to be extremely thin, the etch must be extremely anisotropic, to provide vertical sidewalls. Most non-polymerizing Flourine based etches (SF_6 for example) are not very anisotropic, and therefore not acceptable. Finally, a main constraint of the fin etch is to leave as much of the oxide hardmask as possible, as this will serve as the gate-to-gate isolation structure. For all of these reasons, a Chlorine based reactive ion etch chemistry was chosen. Experiments were conducted on a well characterized Cl_2 etch on the

PlasmaTherm SSL-720 to determine the etch rate, selectivity, anisotropy and oxide erosion. This etch is a three step etch. In the first step, the chamber is cleaned and dehydrated at 40mT with 14sccm BCl_3 and 7sccm H_2 with 200V RF bias for 30 seconds. In the second step, any native oxide that had formed on the exposed silicon surface is etched away at 40mT with 6sccm Cl_2, 42sccm BCl_3 and 21sccm H_2 with 300V RF bias for 30 seconds. The final step of the etch is the deep Silicon RIE at 30mT with 97sccm Cl_2 and 2sccm BCl_3 with 150V RF bias. The final stage of the etch has an etch rate of approximately 88nm/min for single crystal silicon, of ~110nm/min in polysilicon, and 1-3nm/min in thermal oxide. In order to ensure that the etch completed, especially between tightly nested fins, the final stage was run for 1.5 minutes. A clean in Acetone and IPA immediately after the final stage of the RIE was used to remove residual Chlorine from the wafer surface. If not removed, this chlorine can continue etching the fin laterally even after the wafer is removed from the RIE plasma.

Measurements were taken on the etched structure to determine exact film thicknesses and oxide erosion. Reflectance measurements on large features were corroborated with profilometer measurements on the same features to determine nearly the exact final thickness of the SOI layer and the remaining oxide on top of the SOI. Profilometry on the SEM array gave additional information of the extent of the oxide erosion on small linewidth features. Although the profile and shape of the remaining oxide can only be well determined via cross-sectional SEMs, the profilometer served to tell the maximum remaining oxide thickness (i.e. in the center of the hardmask line).

5.3 Gate Stack Deposition

5.3.1 Challenges

The next process module in the development of IG-FinFETs is the growth and deposition of the gate stack. The main challenges posed by this module are due to the inherent three dimensional and non-planar nature of the device. Since the device interface is on an etched surface, great care must be taken to clean and prepare the surface before gate dielectric growth. In comparison with planar devices, the gate electrode deposition is also quite complex. The gate electrode must be deposited in a manner that provides for uniform and conformal coverage over the fin. Finally, engineering the dopant profile in the gate electrode of an IG-FinFET is quite challenging due to the geometry of the device. In planar devices, achieving the proper dopant profile in the gate electrode is relatively simple. Based on the chosen source/drain doping schemes, a polysilicon thickness is chosen that will result in the proper gate dopant profile. This is not possible in the IG-FinFET integration, where the polysilicon thickness at the time of the source/drain implants will be fixed by the fin height. Also, the dopant profile in the lateral dimension (how close the dopants are to the vertical gate oxide) is important in the FinFET, whereas this concern is addressed by the polysilicon thickness in planar devices.

5.3.2 Surface Preparation

Preparing the fin sidewall surface is extremely delicate as, at this point in the process, the fin body is in its most fragile and exposed state. Although cleaning the wafer with Acetone and IPA should remove most of the loose resist and etch polymer, only a sacrificial oxidation and strip will remove particles that have been embedded in the silicon surface by the RIE. This sacrificial oxidation should have two other benefits. By oxidizing the surface, this step should help to smooth out any roughness

in the fin edge resulting from lithography and etch variations. Since this sacrificial layer will be removed, this step will, in effect, remove the outermost silicon from the fin. While the outer surface of the fin is removed, any particles embedded in the surface, and any shallow damage caused by the RIE will be removed as well.

The initial wafers were MOS cleaned without an HF dip to prevent fin cap erosion. These wafers utilized a sacrificial oxidation process of dry oxidation in a TCA ambient at 900C for 10 minutes. According to past data, this should target 90A. Measuring the oxide grown on the sidewalls of the fin is impossible, so monitor wafers were measured. Elipsometer measurements on bulk monitor wafers yielded thickness measurements of 84A. However, this is not necessarily the thickness grown on the sidewalls. Since the wafers are (100) SOI crystal orientation and the fins are etched either in line with the notch or 90° rotated from the notch, then the surface orientation on the sidewalls of the fins is (110). Data suggests that oxidation on this plane occurs at an increased rate [28]. In addition, since this sidewall plane may be rougher than the polished surface of the monitor wafer, the oxidation may grow at a different rate. This sacrificial oxidation can be stripped in the 10:1 HF that is used in the MOS clean. 10:1 HF targets a removal rate of 350A of thermal oxide per minute. A 17 second dip was used to target the removal of 100A of thermal oxide. This should provide enough over-etch to remove any additional oxide that grew on the sidewalls relative to the top surface of the monitor wafer. This step is extremely critical due to the simultaneous etch of the sacrificial oxide and the fin hardmask oxide. Since this integration scheme relies on the existing fin hardmask oxide to isolate the two gates from one another, the removal or erosion of this fin hardmask must be tightly controlled. The initial wafers utilized a 100nm thick hardmask. After removal of 100A (10nm) from the top of the hardmask, 90nm would remain. This would still be sufficient to isolate the gates in CMP. However, the oxide will also etch on the

sides of the hardmask, thinning the hardmask above the fin. If 10nm were removed from either side of a 25nm fin hardmask, only 5nm would remain. This 5nm oxide line would certainly not survive subsequent processing. Results and process modifications will be discussed at the end of the gate stack deposition section.

5.3.3 Well Implantation

Since the device is fully depleted, the threshold voltage is determined primarily by the gate workfunction, oxide thickness and fin thickness, body/well doping has less effect in a fully depleted device than it would in a bulk FET. Also, since the three dimensional volume of the body of a FinFET is so small, random dopant fluctuation in the body will be more severe. In an effort to optimize carrier mobility by reducing scattering, undoped (or insignificantly doped) bodies are preferable. Most wafers were processed without the implementation of a body or well implant. A n-well implant was performed on the final PFET wafer to determine the effect of body doping. This implant was performed at this point in the process, so that the sacrificial oxide would protect the fin from the sputtering effect of the ion implantation beam. Wafers were implanted with $5e^{11}$ cm^{-2} Phosphorus (P^{31}) at 30keV, using a 45º tilt and continuous rotation. Since the geometric profile of the device is different from conventional planar devices, TCAD tools, specifically Silvaco DevEdit® and Athena® Simulation Tools, were employed to determine process conditions. Simulations were run to estimate the final dopant concentration after this N-Well implant, and subsequent thermal processes. Both a 50nm and a 100nm fin were simulated through fin etch, sacrificial oxidation, N-Well implantation, sacrificial oxide removal, and gate oxide growth.

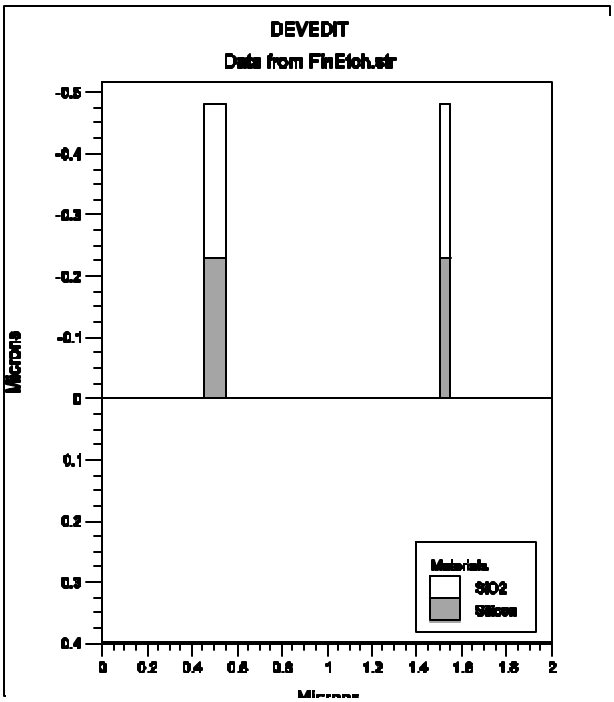

Figure 41. Simulation image of idealized 100nm and 50nm etched silicon fin, with Buried Oxide below and hardmask oxide above.

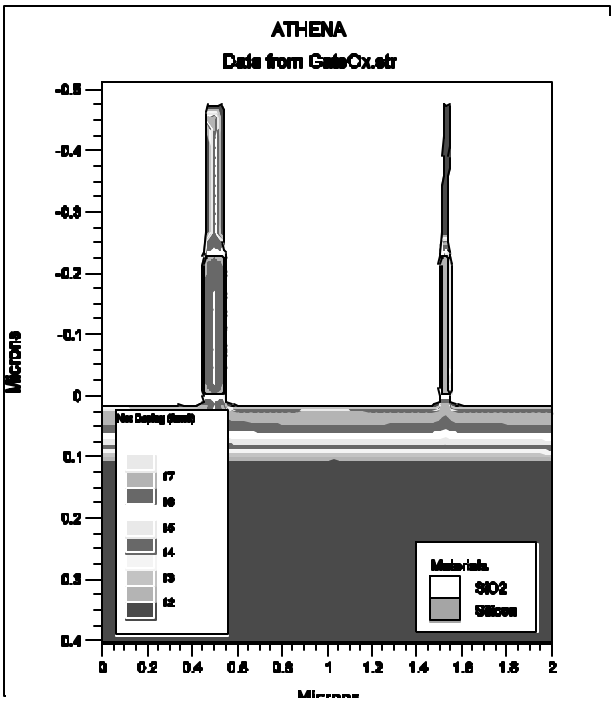

Figure 42. Simulation image showing 100nm and 50nm fin structures after sacrificial oxidation, implant, oxide strip, and gate oxidation.

The profiles of the fin and oxide hardmask are affected by the sacrificial oxidation, strip and gate oxidation. The hardmask is thinned slightly by the BHF in the sacrificial oxide strip, but still appears to be intact in simulation.

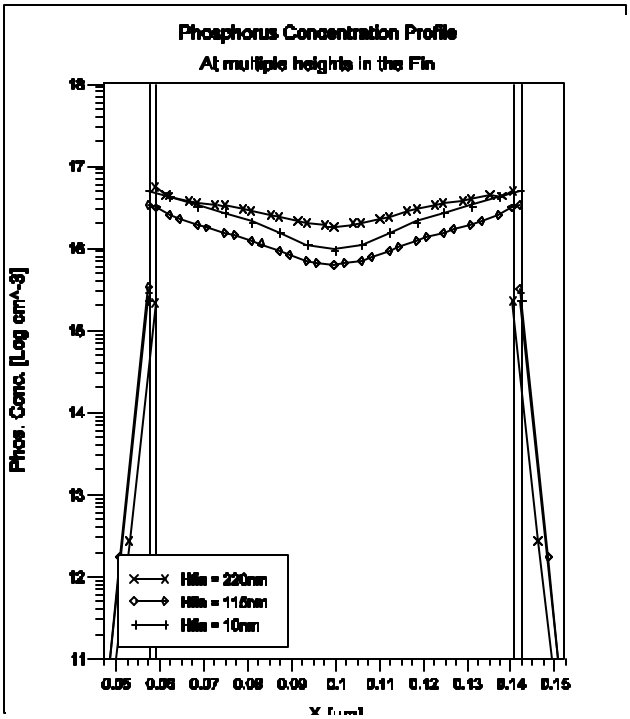

Figure 43. N-Well Implant profile in 100nm fin after gate thermal oxidation at multiple heights within the fin.

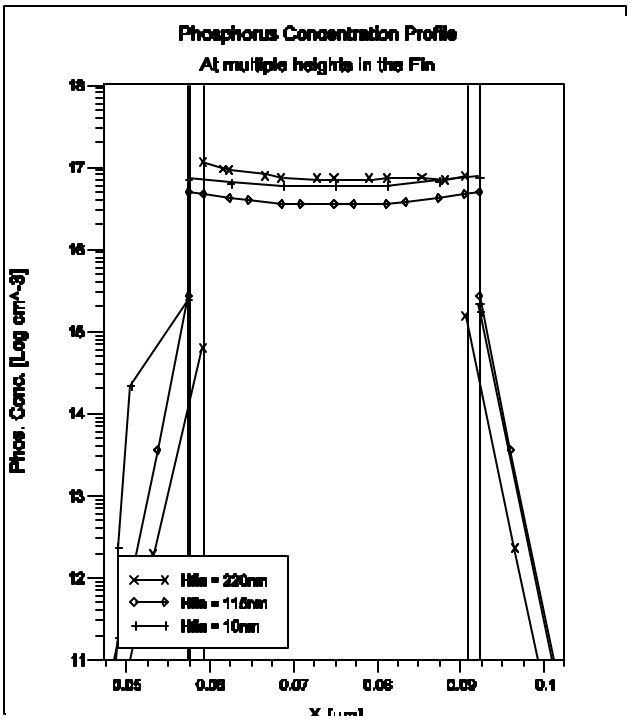

Figure 44. N-Well Implant profile in 50nm fin after gate thermal oxidation at multiple heights within the fin.

The implant profiles from the 100nm fin (Figure 43) and the 50nm fin (Figure 44) show that the high tilt, low energy implant is effective at creating a uniform doping concentration between 5e16cm^{-3} and 1e17cm^{-3} across the thickness and height of the fin. Subsequent thermal processes will help to drive the dopant concentration to an even more uniform distribution.

5.3.4 Gate Oxide Growth

The gate oxide was grown using identical conditions to the sacrificial oxidation. Again, process 20 was used (dry thermal oxidation with TCA) at 900ºC for 10 minutes. Elipsometer measurements on bulk silicon monitor wafers yielded an average of 86.7A with a s = 6.7A across 9 points on the wafer. The monitor that was measured had also undergone the sacrificial oxidation and HF strip. This data suggests that the

17 sec 10:1 HF strip was sufficient to remove the entire sacrificial oxide film from a blank monitor wafer.

5.3.5 Gate Electrode Deposition

The gate electrode deposition was performed immediately after the gate dielectric was grown. As mentioned earlier, a main challenge in the gate stack development is to achieve the desired doping profile in the gate electrodes. Since this integration scheme utilizes a nitride hardmask over the gate to self-align the source and drain implants, the gates can not be doped simultaneously. The gates must either be doped prior to the deposition of the gate hardmask, or after the removal of the hardmask. The initial wafers run through this process were slated to be NFET only wafers. As such, n^+ in situ doped polysilicon can be used. LPCVD polysilicon was deposited using the standard n+ polysilicon recipe, 95sccm of Silane (SiH_4) and 6.6sccm of Phosphine (PH_3) in Nitrogen (N_2) at 300mT and 650°C. This recipe targets a rate of 50A/min. Because the polysilicon will be polished to the height of the existing fin hardmask, the deposition must be at least as thick, in the areas between the fins, as the fins and hardmasks are high. The total height of the fin and hardmask was targeted at 225nm. So, the polysilicon was deposition was run for 60 minutes, targeting a deposition of 300nm. Polysilicon thickness was measured both on monitor wafers and experimental wafers using refractometry. The Leitz film thickness measurement system allows a refractomeric measurement in a small area. Since the areas between the fins contain a relatively simple stack of Silicon substrate, buried oxide and polysilicon, the refractometer measured the poly with high accuracy. Monitor and experimental wafers both showed that the polysilicon deposition resulted in a 375nm film. Since the polysilicon will be polished, this overgrowth is not a strong concern.

Since the intention of the integration scheme studied here is the eventual CMOS integration of IG-FinFETs with conventional three terminal FinFETs, the polysilicon deposition scheme must be changed from the in situ doped method to an undoped deposition and ion implant pre-doping method to accommodate both n+ and p+ doped polysilicon gates. Again, simulations using Silvaco DevEdit® and Athena® Simulation Tools, were employed to determine implant conditions. The simulation consisted of an initial device structure that included the buried oxide, silicon fin and remaining oxide hardmask. A thick polysilicon film is deposited, followed by the gate pre-doping ion implant. This implant simulation was conducted for PFET gate pre-doping evaluation, and therefore Boron was used. After the implant, the majority of the dopant species lies on the top surface of the polysilicon, and far from the device. If the polysilicon is polished before an anneal, most of the gate dopant will be polished away. Therefore, an anneal is required to drive the dopant down towards the gate dielectric. The challenge in this portion of the process development is to achieve adequate gate doping and activation near the gate dielectric, without driving the implant species through the gate dielectric and into the fin. Simulations were run to determine the exact conditions for gate electrode pre-doping implants for PFET wafers.

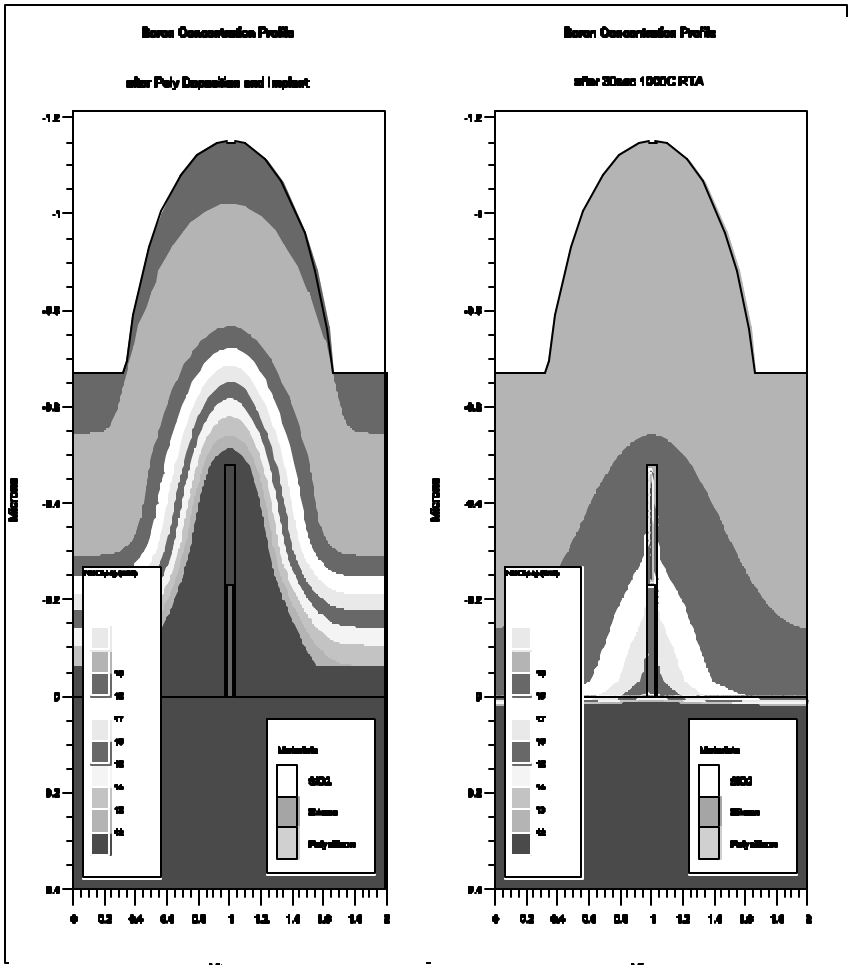

Figure 45. Cross-sectional implant profile simulation results showing Boron concentration in polysilicon gate electrode film after deposition and 1e15cm^{-2} 80keV Boron^{11} implant (left), and after 30 second 1000°C RTA (right).

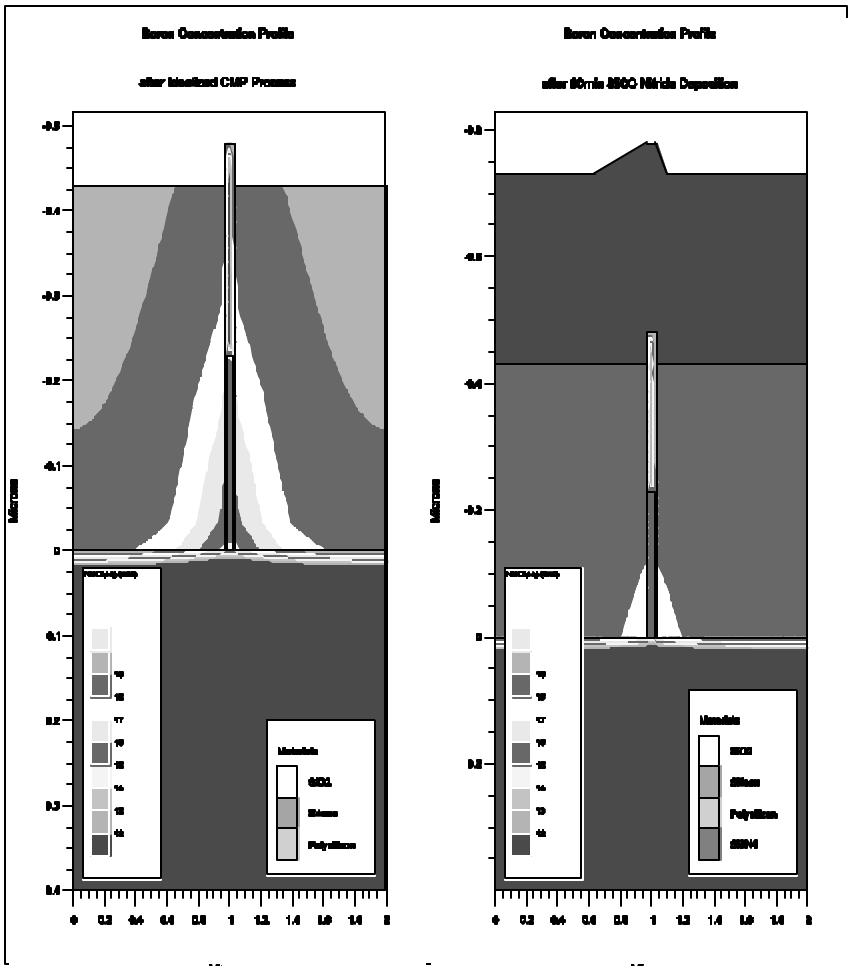

Figure 46. Cross-sectional implant profile simulation results showing Boron concentration in polysilicon gate electrode film afteridealized CMP process (left) and after 60 minute 850ºC Nitride deposition process (used in final integration) (right).

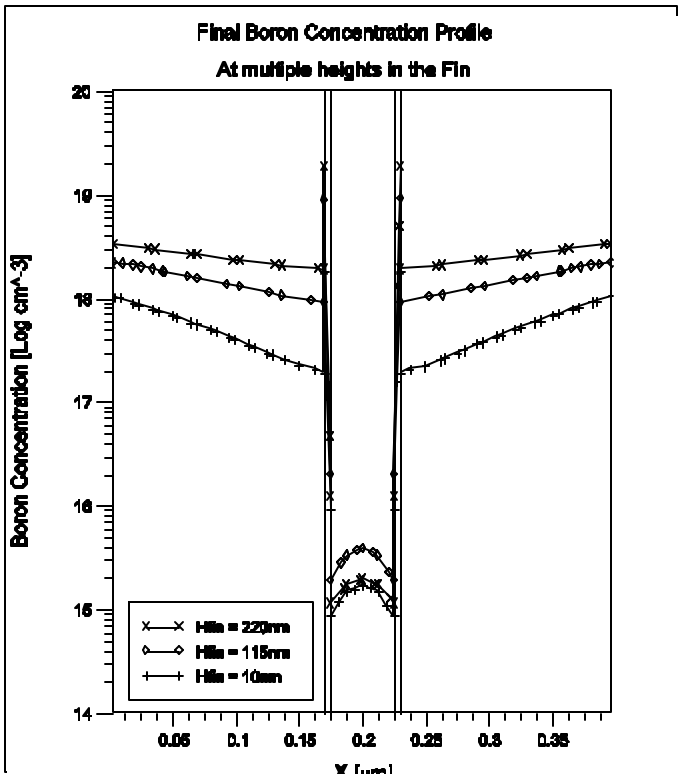

Figure 47. Cross-sectional cut-line analysis of final Boron concentration in gate electrode at multiple heights in the fin

As seen in the cross-sectional implant profile simulation results, the 1e15cm^{-2}, 80keV Boron^{11} implant and 30 second 1000ºC RTA are sufficient to provide a moderately doped, and therefore relatively low resistance, gate electrode after the polishing step. The implant and anneal do not push the implant species too close to the gate oxide and fin, preventing a threshold voltage shift induced by boron penetration. As shown by the cut-line analysis, the gate polysilicon at the very bottom of the fin is doped slightly less than above after a moderate temperature thermal process. Subsequent thermal processes, including the source/drain activation RTA will help to drive the gate doping closer to the gate oxide to reduce the polysilicon depletion effect at the base of the fin. All subsequent wafers were processed using an undoped polysilicon deposition, and gate pre-doping implants as described.

5.4 Gate Electrode Chemical Mechanical Polish

5.4.1 Challenges

The polish of the gate electrode polysilicon represents the most challenging module of the entire integration scheme. The polishing operation is critical to the functionality of the device. The gate polysilicon must be polished far enough so that the remaining oxide hardmask on top of the fin will isolate the two gates. However, if the polysilicon is over-polished below the height of the top of the fin, the resulting transistors will have a portion of the fin un-gated, and leakage consequences will be severe. The CMP selectivity is also extremely critical. The "high" polysilicon, on top of the fin, must be polished away entirely, while the "low" polysilicon, between the fins, must remain. If the "low" polysilicon is polished away, there will be no way to electrically contact the gate of the device. Finally, cross-wafer and cross-chip uniformity is a concern. The CMP must proceed at a consistent rate in all areas in order to yield enough devices to test and characterize for improved device design and process development.

5.4.2 Initial Processing and Results

The initial wafers were polished using a IT1400 Pad and the standard P1000 Polysilicon slurry. The pad was broken in for ten minutes with the in-situ pad conditioner. A recipe was built to run the polish for 30 seconds, with in-situ pad conditioning. The first wafer was polished using 7.5psi of back side pressure for one 30 second cycle. Measurements of polysilicon thickness taken on top of the large rectangular RX shapes showed that polysilicon remained on top of the oxide. Measurements of polysilicon thickness outside of the rectangular region showed that polysilicon remained there as well. After another 30 second cycle with the same CMP recipe, measurements of polysilicon on top of the rectangular structure showed that the

polish had completed, removing the polysilicon from above the remaining oxide hardmask. Optical measurements outside the feature showed that the polysilicon had also been removed. Profilometric measurements confirmed that the polysilicon had been removed from areas far from RX shapes (>5um), but remained in areas very close to RX shapes. This level of dishing is unacceptable for the designs utilized in this set of experiments. A second wafer was polished using 5psi of back side pressure, and results improved. Optical measurements performed after two 30 second polishing cycles showed that polysilicon had been removed from the top of the rectangular feature, but 200nm remained in between features.

After the CMP was completed, a quick recess etch was performed to etch the polysilicon about 15nm further below the point of CMP completion. This is to ensure that the two independent gates are fully isolated. This etch is performed in the SSL-720 using the high-speed silicon etch as used for the fin etch, for only 15 seconds.

Cross sectional SEM images of devices fabricated in this first process sequence were examined to determine the success of the CMP process. Due to the extremely small feature size involved, the cross-sections were taken from the SEM macro, where fins and polysilicon run nearly half the length of each die. As described in Chapter 4, these features are much simpler to cross section due to there length, but are not exactly in the form of the transistors in the device measurement array.

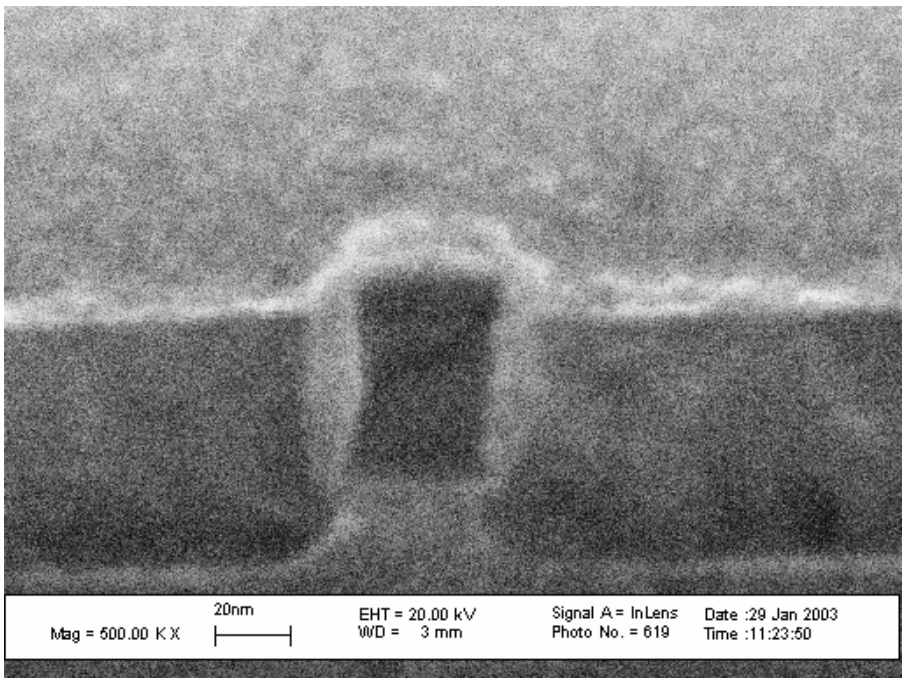

Figure 48. Cross-sectional SEM image showing the fin, remaining hardmask oxide and two independent gate electrodes.

The images show several pieces of process information. First, the gate dielectric appears significantly thicker than the thickness measured on planar bulk monitor wafers, and relatively non-uniform along the height of the fin. This may be a result of the increased growth rate due to either the (110) fin sidewall surface or the rough etched sidewall surface. Or, this may be an artifact of incomplete sacrificial oxide removal. The fin shows a fairly non-uniform cross section, especially on the left side where an extremely concave surface is seen. This profile is probably related to the incomplete removal of the sacrificial oxide, causing more silicon consumption at points where the oxide was removed, and less in points where the sacrificial oxide remained during gate dielectric growth. The increased gate oxide thickness at the middle of the fin height seems to confirm this hypothesis. The non-uniformity in the fin cross-section will cause a variation in threshold voltage at different points along the height, effectively the device width, of the transistor.

The buried oxide below the fin appears to be significantly footed on either side of the fin. This suggests that either the fin etch was not properly optimized to stop on the buried oxide, or that the subsequent wet etches were too aggressive, and removed a large amount of the buried oxide. The oxide hardmask remaining on top of the fin appears to be significantly shorter than expected based on measurements of fin and hardmask after fin etch. A final observation, tightly coupled to the oxide erosion, is that the polysilicon appears to be polished significantly below the height of the top of the fin. The oxide erosion may have occurred in a number of steps, including the wet pre-cleans preceding both thermal oxidations, the sacrificial oxide removal and the polysilicon CMP. The polysilicon polish appears to have proceeded far beyond the intended point. This may have occurred because of the unexpected oxide erosion, of due to lack of control of the polysilicon CMP process. This also suggests that the optical measurement technique used to approximate CMP completion may not have been adequate. Since the buried oxide appears to have been over-etched so significantly, it was not adequate to base CMP success on an optical measurement of polysilicon remaining.

5.4.3 Structural Development

The first major integration change employed was the increase in the height of the fin and hardmask. The SEM results from the first run showed that the fin etched with nearly vertical sidewalls and that the aspect ratio required to etch a 25nm fin through 100nm of SOI (~4:1) is not beyond the capabilities of the lithography and etch processes. However, the small heights of the fin and hardmask increase the required process control of CMP, by limiting the allowable polished height variation. Again, the polish must terminate when the polysilicon height is between the height of the top of the fin and the top of the remaining oxide hardmask. Seeing that the oxide

hardmask in the initial run had been eroded significantly, additional hardmask thickness will increase the process window. By increasing the target thickness of the hardmask to 250nm, the initial SOI thinning oxidation step can be eliminated. This may have the additional benefit of improving cross-wafer fin height uniformity by determining the fin and hardmask thicknesses in one thermal oxidation step. With a single wet oxidation process, the hardmask can be grown to be 250nm, leaving roughly 230nm of silicon remaining from the original 340nm thick film. Since this change increases the final fin and oxide hardmask height, additional polysilicon must be deposited to ensure that the gate electrode, after polish, would exist on the entire height of the fin.

In order to etch the thicker oxide hardmask, a thicker resist is required. NEB-31 can be used undiluted to provide a thicker film. If NEB-31 is spun at 4000rpm for 60seconds, the resulting film is between 240 and 250nm thick. This thicker resist requires a complete re-characterization of the electron beam exposure doses. Initial doses in the e-beam job file had to be adjusted from 105 to 42.5uC/cm^2 for the first write file (fins) and from 17 to 5.5uC/cm^2 for the second write file (large areas). The thicker film required lower exposure dose due to the increased chemical amplification resulting from more photo-sensitive material present. It has been suggested that the use of undiluted resist can improve line edge roughness in very fine features as a result of the local dilution and exposure of non-photo-sensitive material in diluted resists. This will be discussed further in Section 5.5.6.

The second process improvement made for subsequent runs is the decrease in target sacrificial and gate dielectric thicknesses. Electrical measurements taken on wafers produced in this initial process sequence showed negligible gate leakage, indicating that the resulting oxide thickness is well above a thickness where gate leakage becomes a dominant degrading mechanism. These results will be discussed

further in the following chapter. The reduction of the sacrificial and gate oxide thickness will help the process control in a few ways. First, if the oxide is growing thicker on the fin sidewalls that on planar bulk monitor wafers, then a decrease in target thickness (as measured on monitors) will adjust the sidewall oxide thickness to the proper value. This reduction in gate dielectric thickness will improve the device performance. This reduction will also reduce the required time in HF to etch away the sacrificial oxide, and therefore reduce the associated hardmask erosion in that operation. The initial recipe was run at 900ºC for 10 minutes. This is an extremely short amount of time for a thermal process. This short amount of time increases the possibility of non-uniformity in the oxide growth. In order to improve uniformity of sacrificial and gate oxide thicknesses, the recipe was adjusted to use a lower temperature. The sacrificial and gate oxidation recipes for subsequent wafers were changed to 25 minutes at 850ºC, targeting a final thickness of 70A. Elipsometric measurements taken on planar monitor wafers showed thicknesses of 78.9A with a very tight distribution of s = 2.9A across nine points on the wafer.

5.4.4 CMP Development and Results

As discussed in the preceding chapter, fill shapes were added to the design in order to control the local variation of polished polysilicon height. These fill shapes do little to affect the global (wafer level) variations seen in CMP. This type of global variation can be dealt with by process modification and optimization. One major problem with the experiments run for this work is the small amount of area used in the center of each wafer. The six by six array of die are all printed within the center 42mm diameter (to the farthest corner of active area) of a 100mm wafer. The remaining area of the wafer is not patterned, and this has a serious effect on the CMP. Since the e-beam lithography used to pattern the RX level utilizes a negative resist, the

area outside the die that is not exposed is cleared of resist. During the oxide hardmask etch and subsequent fin etch, all of the oxide and silicon are removed from the large area outside the die. This leaves the majority of the wafer at a level that is "low" relative to the features on the patterned die. After the polysilicon is deposited, the only areas on the wafer that has "high" polysilicon to be polished in CMP are the patterned RX features. Since this is such a small proportion of the wafer, it can be polished at extremely high rates, but with very little control. Also, since the majority of the wafer projects "low" polysilicon height before CMP, the wafer may flex and bow on the CMP chuck, further reducing the control of this procedure.

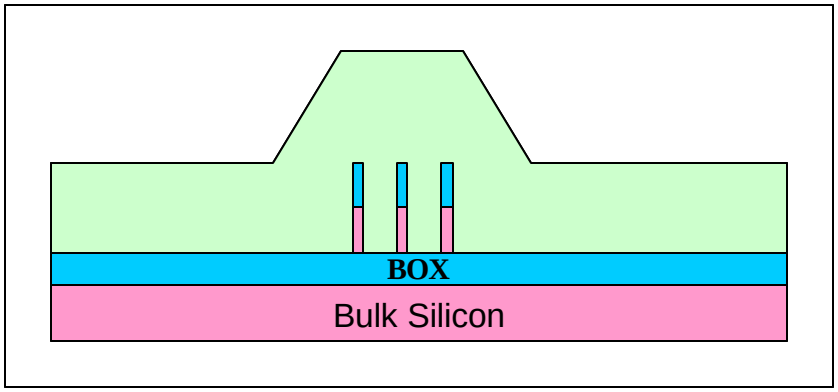

Figure 49. Illustration of wafer cross section showing relative heights of deposited polysilicon, RX features and outer area of wafer

If the outer area of the wafer were exposed during the e-beam lithography operation, then the silicon and the hardmask oxide would remain before polysilicon deposition. This would produce a relative "high" area on the outer regions of the wafer. However, this would increase the e-beam write time to unacceptable levels. However, in order to maintain the oxide and silicon, the exposure must be done at the same time as the RX e-beam exposure. NEB-31 is sensitive to, in addition to electron beam dose, Deep Ultra-Violet (DUV) dose [29]. One of the CNF contact/proximity aligners (the HTG 3HR) utilizes a wideband light source. So, a mask was produced

for use on the HTG that contained only a square of chrome in the center larger than the area of the six by six die matrix by 3mm per edge. This allowed the pre-exposure of the NEB-31 on the HTG, before the e-beam exposure on the VB6 without alignment. After the polysilicon deposition, the wafer cross section would project a majority of the wafer area as "high" polysilicon to the CMP process. This would significantly slow the polish, as much more polysilicon would have to be polished before completion. By slowing down the process, addition control could be maintained on the stopping point. Also, since the majority of the wafer would have oxide hardmask and SOI remaining, a natural "polish-stop" would be presented to the CMP process [30]. When the polysilicon on the outer areas of the wafer completed the polish, the oxide on the majority of the wafer would prevent the polish from proceeding in the critical area in the center of the wafer.

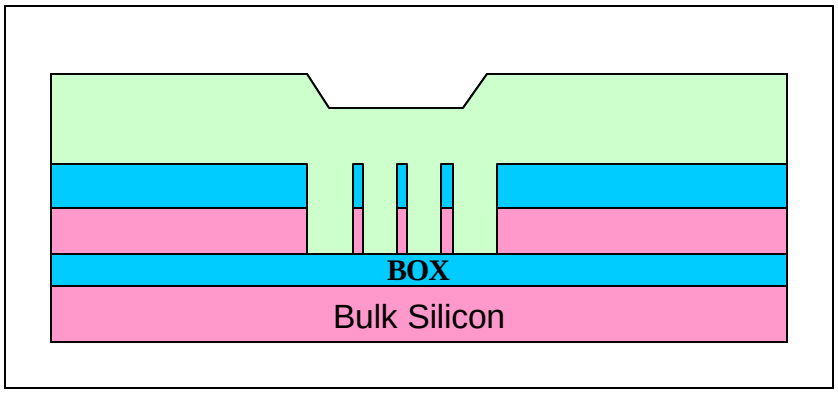

Figure 50. Illustration of wafer cross section showing relative heights of deposited polysilicon, RX features and outer area of wafer with additional pre-exposure.

The only drawback to this process is that a significant amount of time is used at the beginning of the polish to remove polysilicon on the outer areas of the wafer, areas that are non-critical to the device experiments. Since the polish is performed in 30 second intervals with post-CMP cleaning and refractometric measurements done in between each polish step, increasing the polish time to include more 30 second steps

can appreciably increase the overall process time significantly. If the polysilicon in the outer area of the wafer can be removed with a faster process before proceeding into CMP, then the overall CMP process would be significantly shortened. This was achieved by reusing the same RX pre-patterning mask with an opposite polarity resist. A standard Shipley 1818 (positive) resist mask was applied and exposed, again using the HTG contact/proximity aligner. The polysilicon in the exposed regions was etched using the SSL-720 and the high-speed polysilicon etch process, followed by subsequent resist removal RIE and wet cleans. At this point the majority of the wafer projects a "high" silicon and oxide feature to the CMP process, but the polysilicon is only "high" in the critical regions in the center of the wafer. This decreases the overall CMP process time significantly, while still maintaining the natural "polish-stop" on the "high" silicon and oxide features [30].

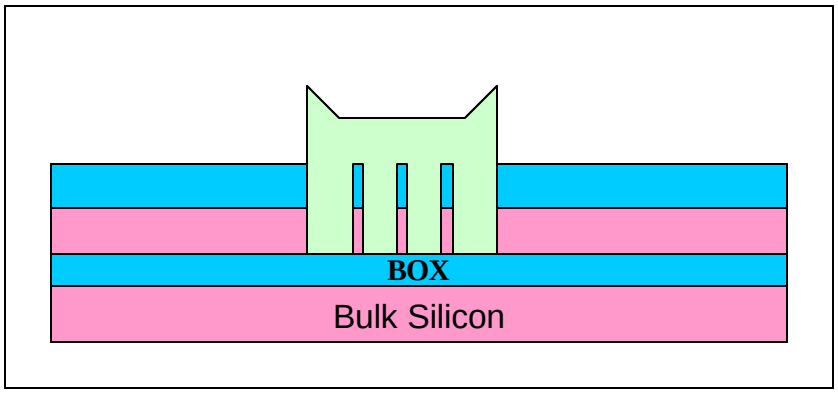

Figure 51. Illustration of wafer cross section showing relative heights of deposited polysilicon, RX features and outer area of wafer with additional pre-exposure and post-deposition non-critical polysilicon removal.

Experiments were run to determine the selectivity, uniformity and overall process time of the CMP process using the pre-exposure and the post-deposition removal. Wafers were patterned through RX, and polysilicon was deposited. The wafers were processed in separate lots, so the starting polysilicon thickness was

different for the three sets of samples (Sample A: No pre-exposure or removal, Sample B: Pre-exposure with no removal, Sample C: Pre-exposure with removal). Refractometric measurements were taken in between each 30 second CMP interval to determine polish completion. Two thicknesses for each polish technique were measured: Remaining "high" polysilicon to be removed, and "low" polysilicon not to be removed.

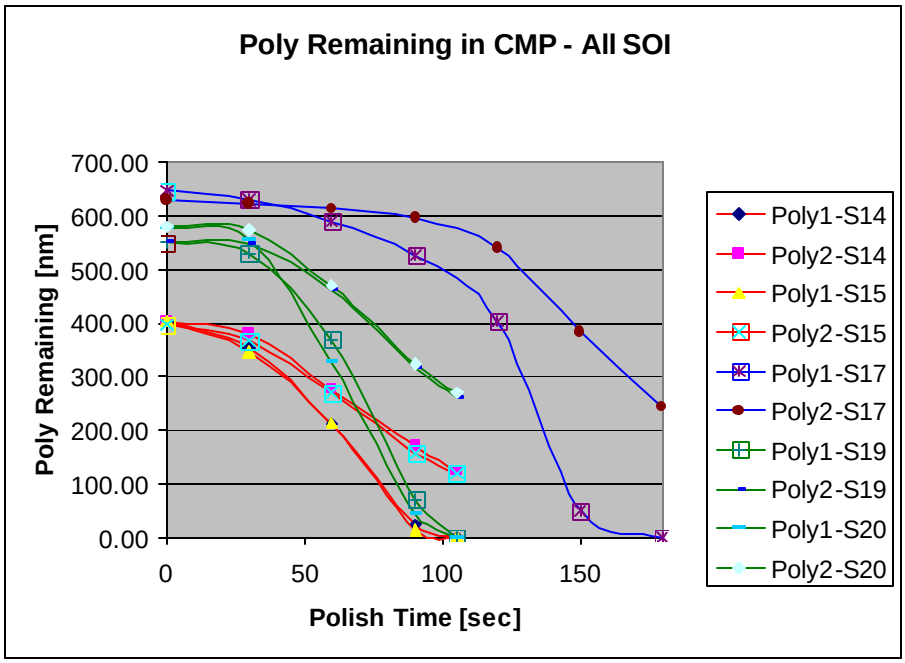

Figure 52. Data from CMP experiment showing thickness of polysilicon on "high" and "low" features used to determine selectivity and process time.

Figure 52 shows the two thickness measurements over process time. "Poly1" represents the thickness of "high" polysilicon yet to be removed. "Poly2" represents the "low" polysilicon remaining between RX features. It is clear that the initial technique (wafers S14, S15) with no pre-exposure or removal had the lowest selectivity (difference between the slope of the Poly1 data and Poly2 data), the lowest remaining polysilicon after the polish. This technique, however, did use a small

amount of process time. The second technique (wafer S17) with pre-exposure but no removal has improved selectivity and leaves a greater amount of "low" polysilicon after completion, but takes at least two more 30 second polish intervals, a significant amount of additional time. The third technique (wafers S19, S20) with pre-exposure and post-deposition removal showed equivalent selectivity and final "low" polysilicon height to the second technique, but occupied significantly less process time. All subsequent wafers were processed using the final technique with RX pre-exposure and post-deposition polysilicon removal.

5.4.5 Results of Process Developments

Cross sectional SEM images verified that the modifications to the height of the fin, the thickness of the hardmask, the target thickness of the sacrificial and gate dielectrics and the CMP process improved the profile of the polysilicon.

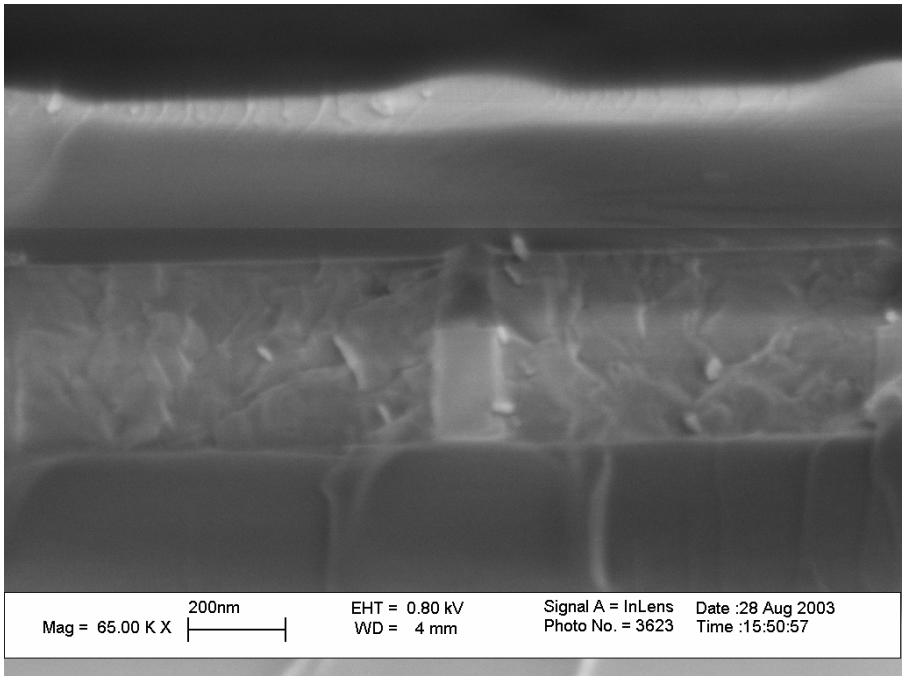

Figure 53. Cross sectional SEM of Fin with remaining hardmask separating polysilicon gate electrodes

The SEM image shows several pieces of information important to the process development, and eventual success of the device design. This fin, while 50nm in the design, is physically closer to 100nm. The fin thickness in the earlier process runs matched the design dimensions more closely. The difference in thickness on this run may be attributed to two main sources of expansion. Since this run used a thicker, undiluted resist to pattern the oxide hardmask, the linewidth may have expanded as a result of the increased chemical amplification due to the increased concentration of photosensitive NEB-31. The oxide hardmask shows a bit of taper in the sidewall angles, meaning the aspect ratio required to etch this hardmask may have exceeded the capabilities of the etch. With the thicker hardmask, any increase in the sidewall angle of the etched oxide hardmask translates into a large increase of the fin linewidth. However, the sidewall angle of the fin silicon appears vertical, meaning that the anisotropy in the silicon etch at this aspect ratio is not degraded. The fin shows an extremely rectangular cross section, which is important for threshold voltage control as discussed in Chapter 2.

The SEM from the earlier run showed significant buried oxide footing, leading to a difficulty in measuring remaining polysilicon thickness after CMP. The optimization and control of fin RIE and wet cleans appears to have fixed the buried oxide footing in this later run. The fin appears fully etched down to the buried oxide, with no remaining silicon foot, and very little oxide over-etch. The modified wet etch recipes that accompanied the change to thinner sacrificial oxide seem to have significantly less impact on the buried oxide over-etch. Although it is difficult to clearly determine from this image, this SEM seems to show that the gate oxide is thinner and more uniform along the height of the fin. This helps to confirm that the sacrificial oxide was stripped completely before gate oxidation, and that the sidewall surface was cleaned more properly to promote uniform gate dielectric growth.

The cross-sectional SEM also shows positive information about the gate isolation strategy. The polysilicon on both sides of the fin appears extraordinarily planar. In addition, the final height of the polysilicon is well above the height of the top of the fin. The oxide hardmask is visibly exposed through the top height of the polished polysilicon. The profile of the oxide mask poking through the polysilicon is transferred into the capping silicon nitride film, and can also be seen in the top interface on the SEM. Based on this SEM, and several others like it, it appears that the integration changes, both structural and changes to the process flow, have succeeded in correcting many of the problems seen in the results of the first fabrication run.

5.5 Gate Definition and Etch

5.5.1 Challenges

Once the polysilicon had been polished to the proper height, the next operation was to deposit a hardmask to pattern the gate electrode. The CMP results play a major role in the success of the gate patterning step. If the CMP was performed properly, as shown in Figure 53, the surface of the wafer would be totally planar, with the top of the oxide hardmask exposed only slightly through the polysilicon, creating minimal non-planarity in the surface. This planarity is crucial for the success of deep submicron gate lithography, as modern lithographic systems have small depths of field. Besides just masking the gate RIE, the gate hardmask plays an additional role in this IG-FinFET integration. Since there is no polysilicon over the top of the fin, the natural self-aligned source/drain implant mask is gone. Before the implants are performed, the remaining hardmask oxide will have to be removed. Therefore, the gate electrode will have to remain in place to block the fin body from the source/drain implants, and to protect the remaining oxide hardmask on top of the fin in the gate region. The fact that the gate hardmask must remain on the gate during source/drain

implantation is what drove the decision to pre-implant the polysilicon. This integration scheme also requires this gate hardmask material to be chosen to allow a RIE to etch the gate polysilicon selective to the hardmask, and to allow an etch (wet or RIE) to remove the remaining hardmask oxide selective to the gate hardmask. For this reason, silicon nitride was chosen as the gate electrode hardmask material instead of the more conventional tetra-ethyl-ortho-silane (TEOS) deposited oxide hardmask. Otherwise, the gate patterning steps represent a few challenges that are somewhat common to conventional CMOS and conventional FinFETs processing. The gate must have extremely vertical sidewalls after the gate is etched. If the sidewalls are angled, the channel length of the FET will vary along the height of the fin, effectively creating a short channel device at the top of the fin, in parallel with a slightly longer channel device at the bottom.

5.5.2 Gate Hardmask Process

Since the gate was already doped, it was desirable to reduce the remaining thermal steps to lower temperatures whenever possible to prevent the penetration of implant species through the gate oxide and into the fin body, potentially affecting threshold voltage and carrier mobility. For this reason, a deposited nitride film was used in the initial process fabrication runs. The GSI PECVD thin film deposition system was run using Undoped Nitride Recipe #5 (400sccm N_2, 40sccm SiH_4, 1900sccm NH_3 at 2.6T, 250W, 400ºC). This resulted in a deposition rate of 1100A/min. Wafers were run for 2 minutes, targeting a final thickness of 220nm. Experiments were run to characterize the selectivity of the planned gate etch (again, the SSL-720 high-speed silicon etch) to nitride. The LPCVD nitride etch rates and selectivity of this process were well known, but no data was available about the rate and selectivity to the GSI deposited nitride. The experiment used a blanket film of

deposited GSI nitride patterned with the gate lithography process to be discussed later. This was done in order to eliminate uncertainty regarding the pattern factor effects of the etch process. The etch rate was found to be 45nm/min as compared with nominal LPCVD etch rate of 18nm/min and a low stress LPCVD nitride etch rate of 22nm/min. The higher etch rate of deposited nitride versus LPCVD nitride is expected due to the mechanical properties of the film [31]. Since the polysilicon thickness to be etched in the first run was roughly 200nm, 2 minutes of polysilicon etching was planned. This would remove 90nm of nitride. A 200nm nitride film was deposited to provide enough remaining nitride after the gate RIE to protect the existing oxide hardmask from the planned BHF oxide removal.

5.5.3 Gate Lithography

Since the PC to RX overlay is extremely critical in this design, as discussed in Chapter 4, Deep Ultraviolet (DUV) lithography was chosen. The Nikon NSR 1505EX has the capability to print 250nm features with 100nm overlay tolerance. This lithography tool requires several alignment features to be added to the RX level, and printed in the electron beam lithography step. In an effort to reduce the non-planarity of the resist film, a Planarizing Anti-Reflective Coating (ARC) was used. The use of ARC will also assist in the exposure quality in the regions where the PC pattern crosses the RX pattern where the change in the reflectivity of the substrate where the oxide is exposed relative to where the polysilicon is exposed could cause exposure differences. DUV Brewer AR2 ARC was spun on wafers at 1050rpm for 60sec. Wafers were then baked at 170ºC for planarization to 150nm thickness. UV-82 DUV resist was spun on wafers at 3000rpm for 20 seconds and then baked at 130ºC for 60 seconds for a resulting thickness of 450nm. Several focus/exposure matrices (FEM) were run to determine the optimal exposure dose and focus setting for this lithographic

process. With the ARC process in place, the resist can be exposed, developed and stripped in Acetone/IPA several times without damaging the ARC or features below. This makes the process more amenable to FEM and lithographic tuning. Wafers were exposed to print the PC pattern in a 12 x 12 matrix, centered on the 6 x 6 RX die matrix. This was done to help balance out the nitride RIE pattern factor loading effect. The UV-82 was post-exposure baked at 140ºC for 60 seconds and developed in 300 MIF (CD-26) for 60seconds. Before pattern transfer etches can proceed, the ARC must be "punched-through". Experiments were run to determine the ARC etch rate in the Applied Materials RIE tool using the O_2 plasma chemistry. A 45 second etch was shown to completely remove the 150nm ARC film.

5.5.4 Hardmask and Gate Etch

Etch characterization was run on two different etch recipes and tools to optimize the nitride hardmask etch. The Applied Materials RIE chamber was run using a CHF_3 etch chemistry (90W RF Power, 30mT base pressure, 30sccm CHF_3). The Oxford PlasmaLab chamber was run using a CHF_3 and O_2 etch chemistry (150W RF Power, 20mT base pressure, 50sccm CHF_3, 5sccm O_2). The CHF_3-only etch appeared to have higher selectivity to resist, however, the CHF_3 in O_2 etch appeared to have a higher nitride etch rate and more uniform results across the wafer. These results are expected as the presence of O_2 during RIE is known to remove polymer etch byproducts that can protect regions from being etched. This same mechanism contributes to the slight removal of resist, lowering the effective selectivity of the etch. Since both the CHF3-only and CHF3 in O2 chemistries etch oxide as well as nitride, any significant over-etch of the nitride gate hardmask would result in removal of the remaining oxide fin hardmask. This oxide loss is unacceptable, since this oxide will be required to protect the fin during the gate etch. As such, careful measurements of

nitride thickness and calibration of etch rates were performed prior to all gate hardmask etched. Also, the chamber was cleaned in an O2 plasma before each etch procedure.

Once the nitride was etched, the resist was removed using an O2 plasma and an Acetone/IPA clean. The gate etch used the high-speed silicon etch on the SSL-720 described in 5.2.5. As described in 5.5.2, this process has a low but significant nitride etch rate, and so any significant over-etch would result in the loss of nitride from the gate hardmask. This nitride loss is unacceptable, since the nitride will later be required to protect the body of the device from source/drain implants. Because of this, careful measurements were performed to determine the amount of polysilicon to be etched, and to calibrate the polysilicon etch rate. It was possible to measure the completion of the etch optically. This process etched a bit faster between the fins than right next to them, so often the etch would reach the buried oxide in between the fins before it completed on the edge of the fin. Any polysilicon remaining on the edge of the fin would short gates together, and possibly produce shorts from gate to fin. After the etch completion was detected optically, wafers were overetched slightly to prevent any shorting. Following the gate etch, wafers were cleaned in Acetone and IPA to prevent further etching of the gate by chlorine residue.

5.5.5 Nitride Removal and Sidewall Reoxidation

After gate etch, the oxide hardmask remaining on the source and drain regions of the fin needed to be removed. Since this process needed to be extremely selective to both silicon and nitride, a wet BHF chemistry was chosen. At this point in the process, it was discovered that the deposited nitride from the GSI tool is extremely sensitive to BHF. The film etched nearly as fast as thermal oxide in BHF, and was becoming too thin to protect the fin body from source/drain implants. This severely

degraded the performance of fabricated devices, as will be discussed later. The integration scheme was changed to utilize an LPCVD nitride instead of the PECVD nitride that was initially chosen to lower the overall thermal budget of the process. This change drove several other changes. First, the gate pre-doping anneal conditions needed to be changed to prevent dopant penetration through the gate oxide. By lowering drive-in RTA time, the implant concentrations were kept further from the gate oxide prior to CMP and nitride deposition. The Boron gate RTA was changed from 1000ºC for 30 seconds to 1000ºC for 10 seconds. Second, a major change to the metallization process was required. This will be discussed later in this chapter. The LPCVD nitride was significantly more resistant to BHF. Nearly zero nitride etch rate was measured on a monitor wafer.

After oxide removal, wafers were run through a thermal oxidation identical to the sacrificial oxidation and gate oxidation processes. This oxidation served two purposes. First, it would further assist in the elimination of gate shorts by oxidizing any small particles of polysilicon remaining after the gate RIE. Second, this oxidation capped and protected the fin silicon from the sputtering effects of the ion implant beam. Since the fin is so thin and fragile, the addition of an oxide film assists in its stability as well. This oxidation is important to control for the following reason. This oxidation thins only the portion of the fin that is exposed outside the gate electrode. Therefore, the high resistance source/drain regions become thinner, and therefore higher in resistance. If this oxidation is not tightly controlled, the source/drain regions may oxidize completely, and result in open circuits, or the resistance will increase to the point where the device performance will degrade significantly.

5.5.6 Gate Definition Results and Analysis

Top down SEM images taken of devices in several stages of the development indicate process improvements and challenges for future research.

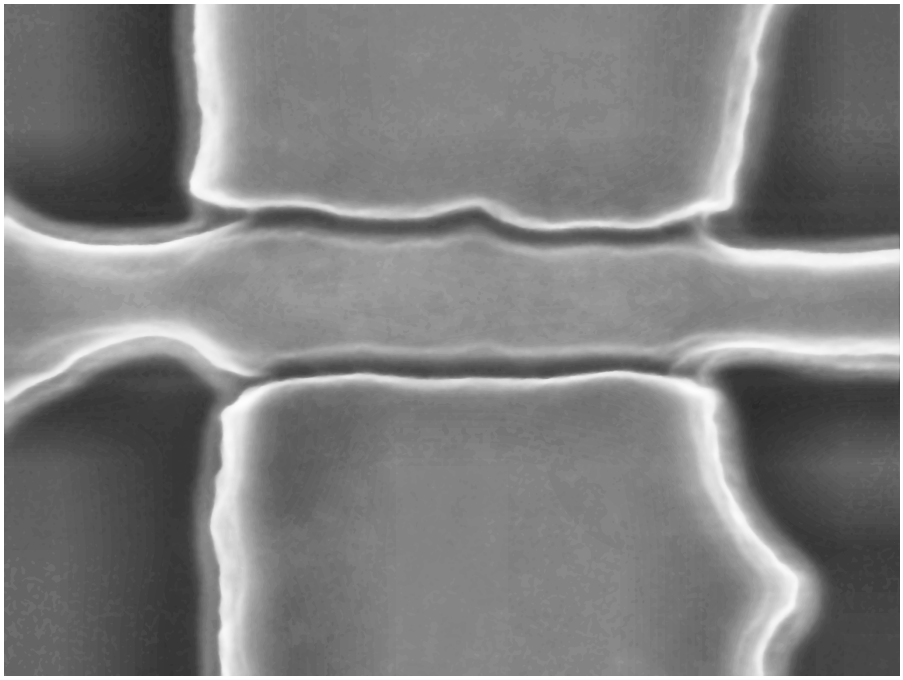

Figure 54. Top down SEM of 50nm fin running from right to left with 250nm gates on top and bottom from initial fabrication process

The image taken on the device from the initial fabrication process (Figure 54) shows a few important details. First, the line edge roughness of the fin (right to left) is quite severe. This may be due to the use of the diluted resist, or the incomplete sacrificial oxide removal. The fin is thinned outside the gate region by the sidewall reoxidation. The source and drain extension regions are clearly thinner than the fin in the body region, but have not been thinned to the point of removal. Finally, this SEM was taken after fin oxide removal. It is apparent that in this run, the nitride gate hardmask was removed by the BHF, because the gate dielectric spacing can be seen.

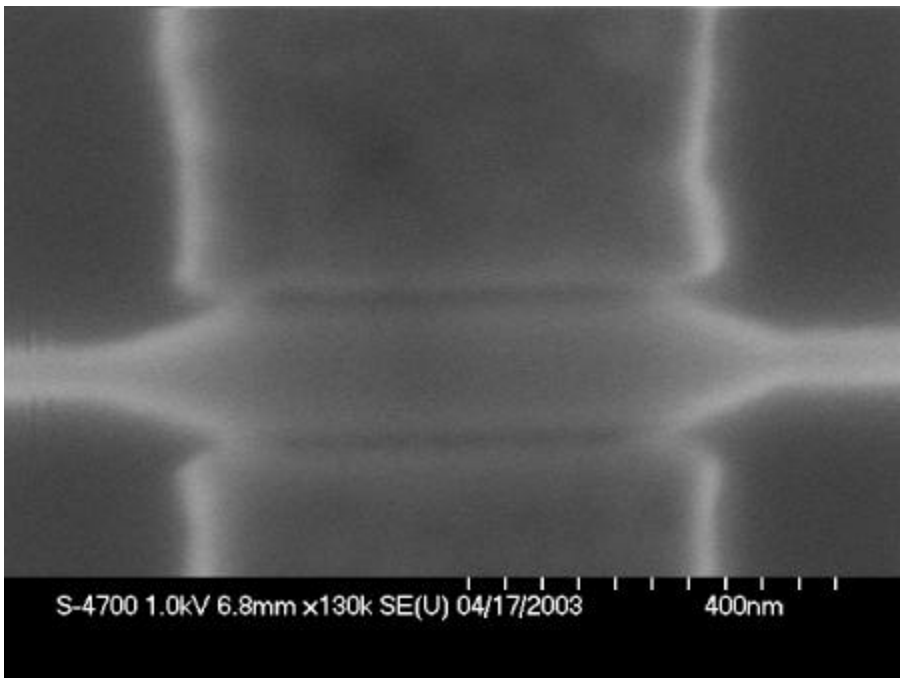

Figure 55. Top Down SEM of 100nm fin running right to left with 500nm gates on top and bottom from later process run.

In later process run, the line edge roughness on the fin was significantly improved due to the change to undiluted e-beam resist and more complete removal of the sacrificial oxide. As seen in Figure 55, loss of control during the post-gate etches and sidewall reoxidation can thin the source and drain extension regions to the point of extremely high resistance. After the change to LPCVD nitride was made, the gate dielectric spaces are no longer visible in top down SEMs, as seen in Figure 56. This is because enough nitride remains after the gate etch and oxide hardmask removal to cover the fin body and the gate dielectrics. This coverage is essential to maintaining an undoped body and proper performance in the FET.

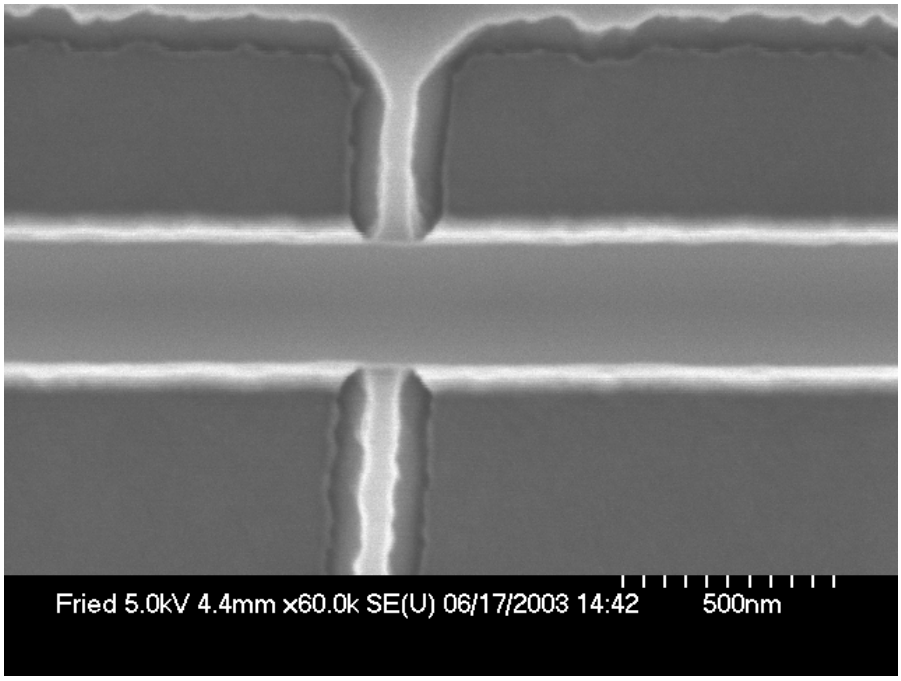

Figure 56. Top Down SEM showing 75nm fin running from top to bottom and 500nm gates on the left and right from the final process fabrication run

Line edge roughness on the fin is still obviously a problem, but substantially improved from the early process runs. Line edge roughness on the gate nitride hardmask and polysilicon electrode appears to be extremely low. This may be due to the improved mechanical stability of the LPCVD nitride during the hardmask RIE, gate RIE and subsequent wet etches and cleans.

5.6 Source/Drain Implantation and Rapid Thermal Anneal

5.6.1 Challenges

The challenges associated with the ion implantation and RTA process development for this device stem from the inherent three dimensional nature of the device. Implant tilt angles and rotations must be carefully chosen to achieve the proper overlap between the source/drain regions and the gates. While a wealth of knowledge and simulation data exists regarding planar CMOS implants, very little

data exists as the process conditions for FinFET implants. Three-dimensional process simulators exist, but were not available at the time of this research. Process development for ion implant conditions was therefore limited to the two-dimensional Silvaco Athena® simulation tool. Modeling inherently three dimensional implant profiles in a two-dimensional simulator presented many challenges.

5.6.2 Implant Process Simulations

Several aspects of the final source/drain implant profile were investigated in simulation. First, since this process integration did not use extension implants or spacers, the source/drain implants must be appropriately overlapped by the gate. This required that the implants be performed at a tilt and multiple rotations. The rotation angle of FinFET implants is essentially equivalent in purpose to the tilt angle in planar CMOS extension/halo implants. For simplicity and symmetry of devices, implants were all modeled and performed at quad angles, 45º offset from the notch. This allows devices oriented both in the X-Fin and Y-Fin direction to be implanted equivalently.

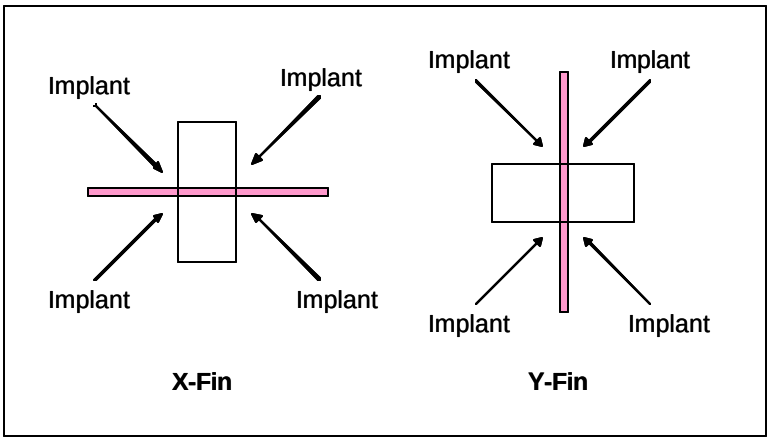

Figure 57. 45º implant rotations ensure equivalent implants for both X-Fin and Y-Fin oriented FinFETs

Second, the source/drain implants had to provide high enough dopant concentration in the fin and large source/drain contact pads to ensure ohmic contacts and low extrinsic resistance to the device. The third requirement of the source/drain implants was to provide a uniform dopant profile along the height of the fin. This was the most difficult requirement and the most critical to the device performance. If the profile is not uniform from the top to the bottom of the fin, then the resulting device will have a non-uniform channel length across the effective width. This is unacceptable, especially at short channel lengths, where short channel effects will cause the threshold voltage to fluctuate with the channel length.

In order to satisfy all three of these implant requirements, several different structures were used in simulations, and the implant conditions were modified to fit the structure needed by the simulation. To determine the eventual overlap distance, a thin-body planar FET structure was used. The body thickness of the planar FET structure was set to be half of the fin thickness of the FinFET under simulation. The implant conditions were adjusted as follows. The tilt in simulation was set to 45º to simulate the rotation of the actual implant. The rotation of the implant was set to 78º to simulate a 12º tilt of the actual implant. Since the simulator uses the simulation plane as the 0º reference for implants, the supplement (90-Angle) of the actual implant tilt angle must be used for the simulated implant rotation. The actual source/drain profile relative to the gate edge can be approximated by mirroring the structure about the bottom of the body.

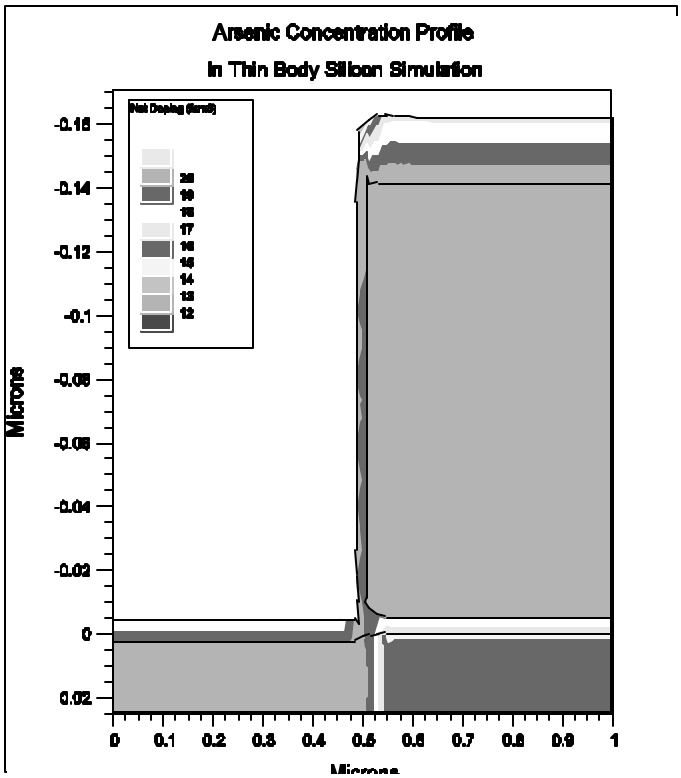

Figure 58. Cross-sectional Arsenic concentration profile from simulation results using thin body (25nm body thickness before sidewall reoxidation) planar FET and angle-adjusted 40keV implants after sidewall reoxidation and implant

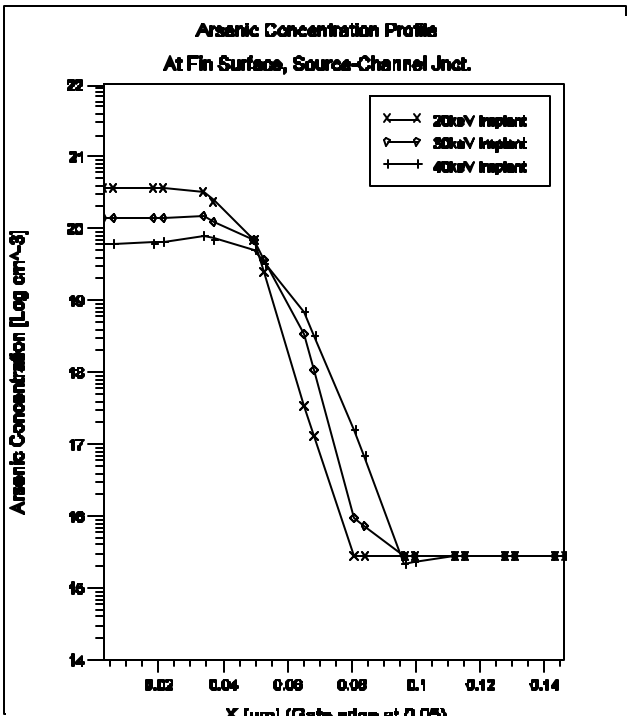

Figure 59. Cut-line analysis of Arsenic concentration profile for multiple energies of implant at the surface of the fin (5nm deep)

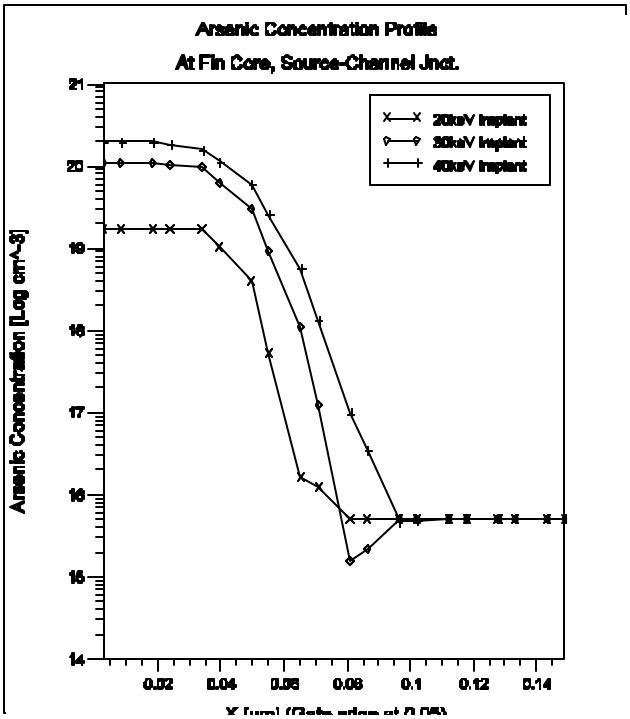

Figure 60. Cut-line analysis of Arsenic concentration profile for multiple energies of implant at the core of the fin (24nm deep)

This method of simulation ignores the implant that approaches the fin from directly above, but is a good method for examining the expected overlap profile at lower heights on the fin. The higher energy implants result in higher dopant concentrations further under the gate, as expected. The 40keV implant shows a full 20nm overlap (Figure 59). The lower energy implants show a substantially lower dopant concentration deep in the fin, also as expected. The 20keV implant shows an order of magnitude lower dopant concentration in the source/drain region deep in the fin (Figure 60). This will result in a substantially increased extrinsic resistance for devices implanted at 20keV. As such, NFET source/drain implants should be run at least at 30keV, more preferably at 40keV. Any higher energy for the implant will result in much larger overlap, and increased short channel effects.

To satisfy the second implant requirement, high dopant concentrations in the contact and extension regions, simple top-down simulations were run. These simulations used structures similar to the fin extension areas. Large and small cross-section features of silicon were implanted in simulation with conditions identical to the actual implants. The results show the doping profile as a result of only the implant species that approaches the fin from the top. This analysis ignores the implant species that approaches the fin from the side due to the tilted, rotated implants. However, it is a useful method for estimating the effect of the top implant species both at the gate edge and in the large contact areas.

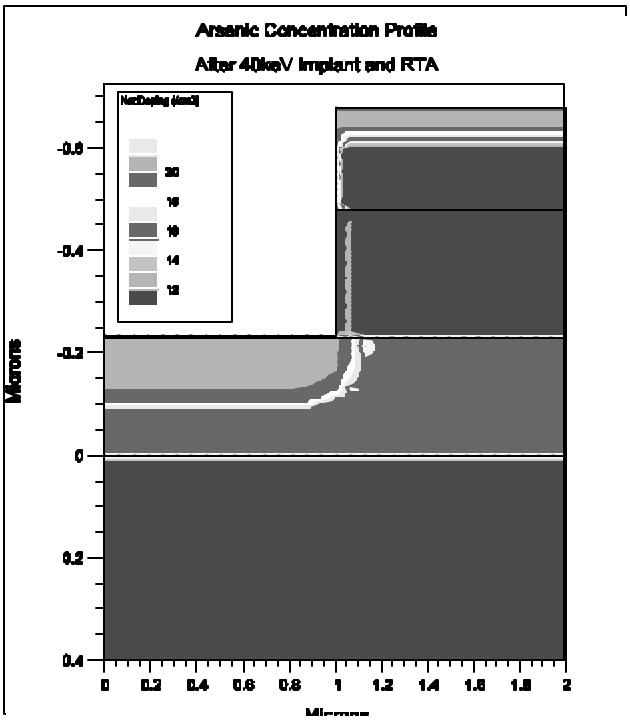

Figure 61. Cross-sectional Arsenic concentration profile in top-down simulation using a thick SOI (230nm) with remaining hardmask oxide and nitride on right half, after 40keV implant ant RTA

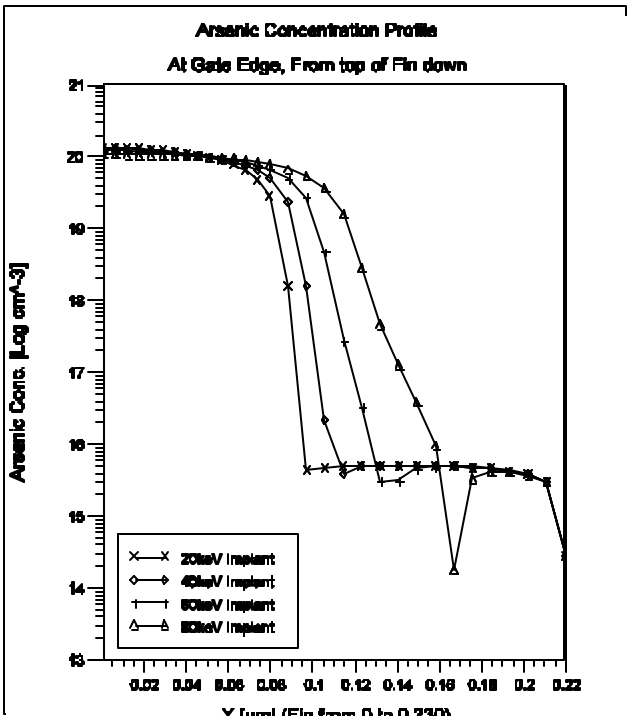

Figure 62. Top-Down Arsenic concentration profile at the gate edge showing depth of implant as a function of implant energy

It is clear and expected that the higher energy implants will have deeper profiles (Figure 62), and therefore lower extrinsic source/drain resistance between the contact and the gate edge. None of the implants simulated saturate the SOI layer completely. However, the higher energy implants are also expected to increase the overlap at the top of the fin the most. The lowest energy implant simulated (20keV) appears to have the most overlap. This is because more of the implant species remains at the top of the fin and can diffuse further. The 40keV implant appears to have a steep profile (Figure 63), and no larger overlap than expected by the rotated simulations discussed above.

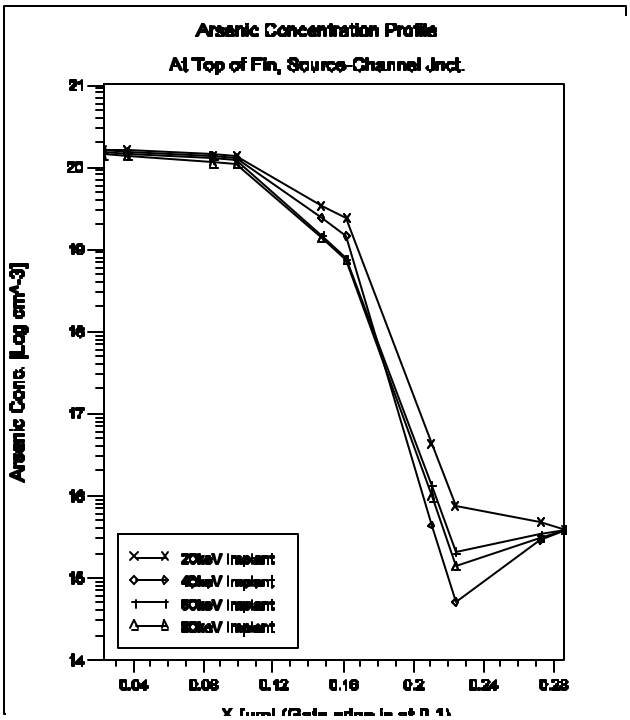

Figure 63. Arsenic concentration profile in the top of the fin under the gate edge for various implant energies

Based on these simulations, the NFET source/drain implant was chosen at 40keV, 8e14cm^{-2}, 12º tilt and 4 quad rotations of 45º, 135º, 225º, and 315º. Based on an identical analysis, the PFET source/drain implant was chosen to be BF_2 at 25keV, 5e15cm-2, 12º tilt and quad rotations of 45º, 135º, 225º, and 315º.

5.6.3 Ion Implantation and Rapid Thermal Anneal Processing

The NFET implants were executed as chosen by simulations discussed above. After the discovery of the GSI nitride etch sensitivity. The final NFET wafers were implanted at a reduced energy due in an attempt to dope the source and drain without implanting dopant through the thin remaining nitride and into the fin. PFET source/drain implants were run with continuous rotation instead of quad angles as a cost savings measure. The profiles are expected to be similar.

Table 4. Implant conditions as processed

Run	*Polarity*	*Species*	*Energy*	*Dose*	*Tilt*	*Rotation*
NFET1b	NFET	As	40keV	8e14	12º	Quads + 45º
NFET2a	NFET	As	10keV	8e14	12º	Quads + 45º
PFET2c1	PFET	BF2	25keV	5e15	12º	Continuous
PFET2c2	PFET	BF2	25keV	5e15	12º	Continuous

Following the implants all wafers were annealed in the AG Heatpulse RTA with the conditions used in the simulations.

Table 5. Rapid Thermal Anneal Conditions as processed (* reflects mis-processing of RTA for run PFET2c1)

Run	*Ramp Up Rate*	*Max Temp*	*Time at Max*	*Ramp Down Rate*
NFET1b	50C/sec	1050ºC	30sec	-50C/s
NFET2a	50C/sec	1050ºC	30sec	-50C/s
PFET2c1	50C/sec	1050ºC *	30sec *	-50C/s
PFET2c2	50C/sec	950ºC	5sec	-50C/s

5.7 Metallization

The process flow was completed with a simple one-level Aluminum BEOL process. The purpose of the metallization was to passivate the fragile devices and provide for probing contacts to the device terminals. This process began with a 500nm undoped oxide PECVD using the GSI system. Contact holes were patterned using the GCA 5X G-Line Lithography system and Shipley 1827 resist, and etched using the Applied Materials RIE CHF_3 plasma chemistry. In the initial integration,

with the PECVD nitride gate mask, the contact hole etch was done in one step for both contacts to polysilicon and contacts to active silicon regions. The CHF_3 chemistry could etch through the BEOL oxide and the thin remaining PECVD nitride over the polysilicon regions while only mildly over-etching the BEOL oxide over the active silicon regions. After the conversion to LPCVD nitride, the contact patterning and etching had to be decoupled for contacts to polysilicon and contacts to active regions. The over-etch of the oxide that would be required to etch through the thick remaining LPCVD nitride would be too substantial to keep the etches coupled. After the contact holes were etched, 500nm of Aluminum is sputtered to give good step conformality. The M1 pattern was exposed using the G-Line Lithography system, Shipley 1827 resist and the YES NH_3 image reversal technique. M1 etch was performed in a Transene Type A wet chemistry and also the Plasma-Therm SSL-720 Cl_2 plasma RIE. Since the critical dimensions of the M1 pattern are not terribly small, both of these methods were essentially equivalent. A final passivation oxide was deposited to prevent probes from scratching and damaging M1 lines. Probe holes were opened in this final oxide using the same lithography and etch scheme as the contact holes. A final 20 minute anneal in 15% Hydrogen was used to terminate surface states in the devices and to promote ohmic contacts between the metal and the doped silicon and polysilicon regions.

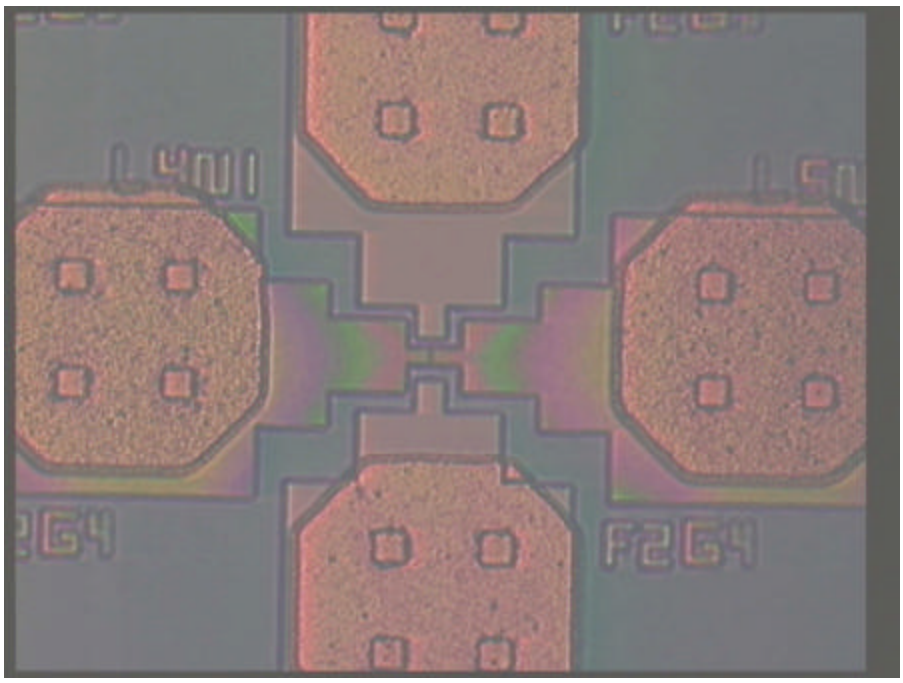

Figure 64. Top-down micrograph showing etched Aluminum probe pads contacting, via visible contact holes, the gate polysilicon (on right and left) and the single crystal silicon source/drain regions (on top and bottom)

5.8 Conclusion

For the first time, a full integrated process has been demonstrated for the fabrication of functional N-type and P-type independent-gate FinFETs. Substantial effort was applied to the successful development of a FinFET process in CNF, targeting fin thicknesses ranging from 10 to 100nm, with gate lengths ranging from 0.25 to 5um. The final process utilizes mixed electron-beam and optical lithography to provide small fin and gate features with tight PC-RX overlay tolerances. The challenges of gate separation and subsequent implant masking have been surmounted by detailed CMP process development and a novel LPCVD Silicon Nitride gate hardmask approach. Early steps toward full CMOS integration, such as a poly gate electrode pre-doping scheme, have been investigated with successful results, as will be discussed in the following chapter. Many of the remaining challenges to full CMOS

integration of IG-FinFETs are common to integration challenges of conventional FinFETs, including optimization of implant profiles, a robust source/drain spacer structure, selective silicon epitaxy for raised source/drain, Silicide formation on thin fins, and a thin fin metal contact architecture.

Chapter 6: Independent-Gate FinFET Electrical Characterization

6.1 Overview

This chapter describes the electrical characterization of final complete IG-FinFET experiments. Results from early process runs are discussed briefly, in order to understand causes and effects of major changes made in the fabrication process discussed in the last chapter. This chapter will primarily focus on the electrical characterization of the most successful process runs, NFET2b and PFET2c2. A challenge to the electrical characterization of the IG-FinFET is the fact that new metrics are required to gauge the quality of a device with two gate terminals. Attempts are made in this chapter to formulate new metrics to compare independent double gate devices. In addition to independent gate measurements, the IG-FinFET is characterized in double-gate mode, with both gate terminals biased together. This provides one way to measure and compare the quality and characteristics of the channel of the device to other MOS structures. Detailed analysis will be presented comparing the different device geometries included in the testsite in both double-gate and independent-gate modes.

6.2 Test Methodology

All DC electrical characterization was conducted using a HP 4156B Semiconductor Parameter Analyzer connected to a simple Cascade MicroTech micro-

manipulator probe-station. Since all tested devices consisted of a single fin, the effective device width, and therefore the saturated drain current, was relatively small. These small device currents did not show oscillations when tested with non-terminated test lines, and so 50 Ohm terminations were not required. To evaluate simple functionality of the devices under test, multiple I_D vs. V_G curves were generated. In general, one gate (called the "front gate" for simplicity) is swept through the range of gate bias with the other gate (called the "back gate") is held constant. The sweep is repeated for various back-gate voltage settings and drain-source voltage settings (typically, the source is grounded and serves as the common voltage reference for the device while the drain voltage is adjusted). This simple test allows the simultaneous evaluation of several structural characteristics of the device. First, the drain current curves, as with single gate CMOS devices, reflect the integrity of the MOS structure and the continuity of the source/drain contacts and extensions. Gate shorts, source/drain shorts and source/drain opens (typically broken fins or improperly contacted source/drain regions in the case of the FinFET) are easily detected with this test methodology. Also, by measuring gate current, the gate-to-gate isolation of the IG-FinFET can be measured. Finally, by comparing multiple sweeps with different back-gate bias, the double-gate nature of the device can be evaluated. To first order, the effect of the back-gate bias on the IG-FinFET should appear as a threshold voltage shift at certain bias conditions and as an additional current source at increased back-gate bias conditions. The following sections will discuss the results of this type of measurement and analysis, as well as additional tests when appropriate.

6.3 Results of NFET1b

The initial testing of devices from the first exploratory process run to completion appeared to show nearly complete shorting from source to drain on every

device. After many devices were tested, a few showed a very small amount of MOS-type current between source and drain.

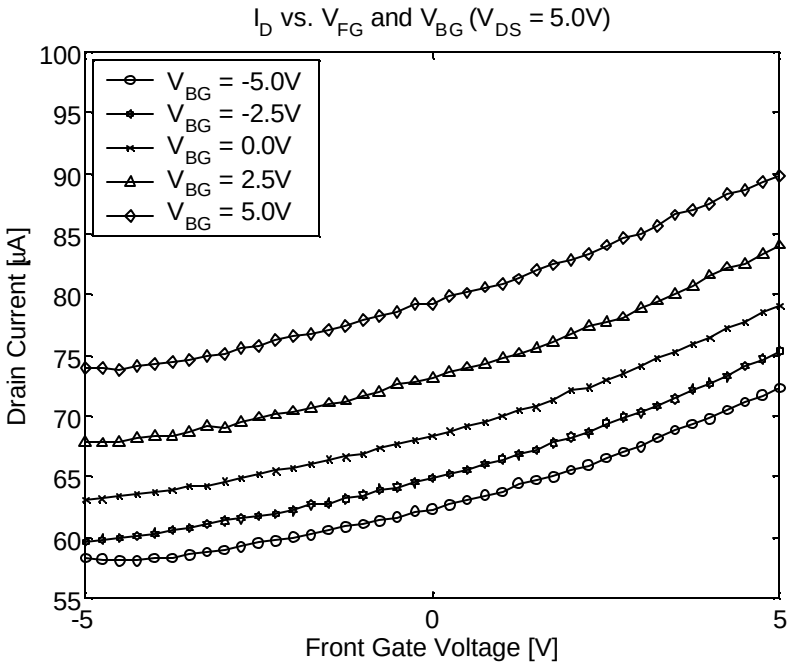

Figure 65. Example of I_D vs. V_{FG}/V_{BG} data from NFET1b process run

These devices showed that a significant shorting mechanism existed between source and drain, but that the MOS structure was possibly intact, although of extremely poor quality. The large amount of drain current at low front and back gate biases (approximately 58uA) represents the inability to turn the device off. The small amount of current increase as the gate biases are increased shows that only a small portion of the device is behaving as a transistor. Physical failure analysis, as shown in Section 5.4.2, confirmed the electrical analysis. The oxide hardmask had been substantially eroded due to several processes, and the source/drain implants were allowed to penetrate deep into the body of the device. This essentially created a resistor in a large portion of the fin, instead of a transistor, leading to the large leakage current, and inability to turn the device off. The combination of this electrical data

with the physical analysis led to several process changes detailed in the previous chapter. These results represented a significant step towards the successful fabrication of Independent-Gate FinFETs [32].

6.4 Results of NFET2a

After several fabrication experiment and resulting integrated process changes were implemented, another full integrated NFET process run was completed. The electrical characterization of the devices from this run showed moderately improved yield, and substantially improved device properties. These devices showed serious punch-through when tested in saturation mode (V_{DS}>1.0V), even for long channel lengths. This indicated that, although the complete source/drain short had been eliminated, there were still some issues with the source/drain junctions. This behavior suggests that the device was behaving like an extremely short channel device in some small portion of the effective width. One possible mechanism to allow this is extremely non-uniform source/drain junction profiles. Improper or marginal implant masking during the source/drain implants could cause the scenario shown in Figure 66. If the gate electrode material were maintained and the gate edges still acted to mask the implant from the sides of the fin, the proper source/drain profile could be produced for a large portion of the effective device width. However, if the implant mask were eroded, the source/drain junctions at the top of the fin would be substantially altered. Implant species would penetrate into the fin much deeper than intended, creating an extremely short channel device at the top of the fin as seen in Figure 66. Physical failure analysis of the eroded PECVD Silicon Nitride implant mask seemed to confirm this hypothesis. Further experiments showed that the PECVD nitride would not be effective as both a gate etch mask and a source/drain implant mask due to its limited dry and wet etch resistance. The combination of this electrical data and physical

failure analysis led to the change from PECVD Silicon Nitride to LPCVD Silicon Nitride in an effort to maintain a proper implant mask.

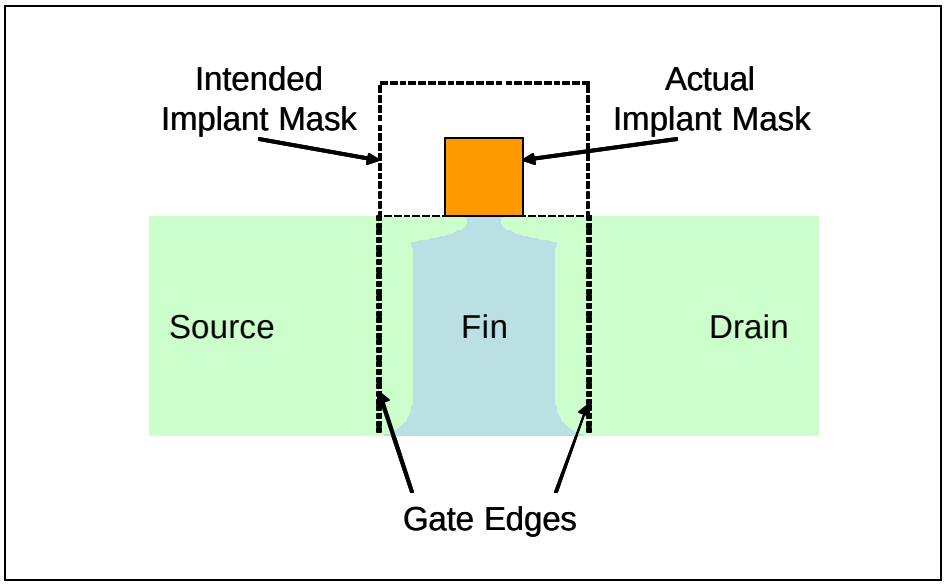

Figure 66. Side view of hypothesized structure of reduced implant mask, and resulting source/drain implant profiles

Due to this effect, testing was conducted only in linear mode, with V_{DS}=50mV. Linear mode I_D vs. V_{FG}/V_{BG} curves, as seen in Figure 67, showed clearly that the independent-gate FinFET structure had been successfully fabricated. I_{OFF} was six decades below I_{ON}, indicating that the source/drain implant had been prevented from doping the majority of the body of the fin, as had occurred in the NFET1b run.

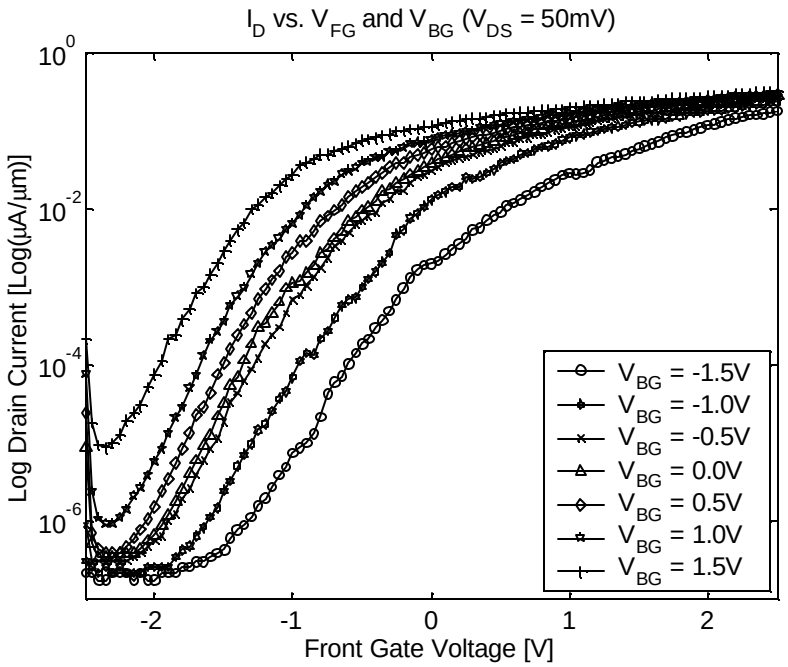

Figure 67. I_D vs. V_{FG}/V_{BG} at VDS = 50mV (linear mode) for T_{FIN} = 50nm, L_{GATE} = 2um, Gamma = 2um device from NFET2a process run. Current is normalized to $W_{EFF} = 2 \times H_{FIN}$.

The device current was lower then expected from simulation, indicating the possibility of several device shortcomings. First, as seen in top down SEM images, such as Figure 55, the source/drain regions of devices in this run were extremely thin, severely increasing the extrinsic resistance. Also, the sacrificial oxide removal wet etch was quite conservative (short time) on this run to try to preserve the oxide hardmask. If the sacrificial oxide were not entirely stripped, the gate oxide interface may contain the RIE polymers that the sacrificial oxide was meant to remove.

Although the on current was low, the threshold voltage was clearly modulated by the back gate bias. At extremely low back gate bias ($V_{BG} \sim -1.5V$), when the body is not fully depleted, the subthreshold slope is degraded due to the additional depletion capacitance. At moderate biases ($-0.5V < V_{BG} < 0.5V$), the back gate alters the potential profile across the thickness of the fin, effectively altering the threshold voltage of the front channel. Finally, at higher bias ($V_{BG} \sim 1.5V$), the back channel

begins to enter inversion, and the additional current of the back channel adds to the aggregate device current. This has two effects on the shape of the drain current curve. First, the additional back channel current elevates the aggregate off current. The back channel current also adds to the current in the subthreshold region, effectively decreasing the extracted threshold voltage further.

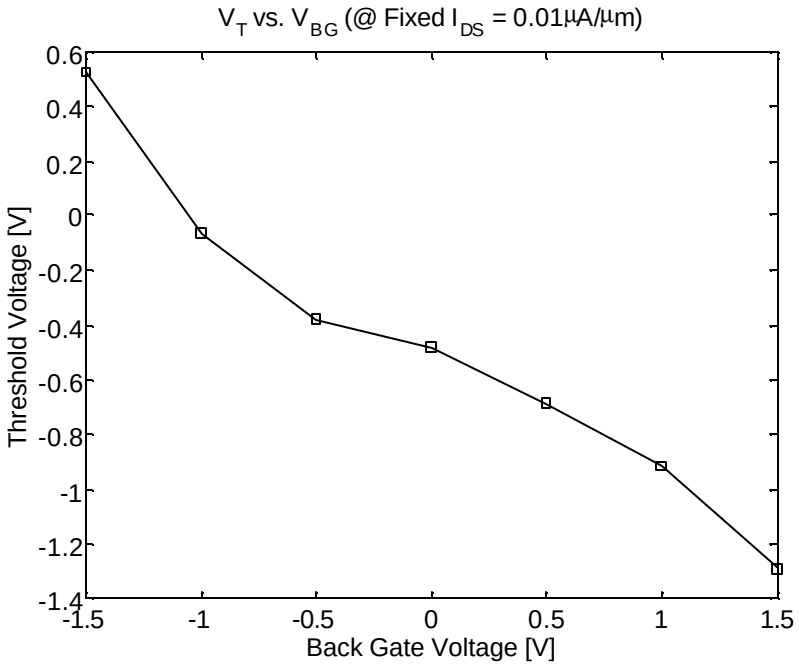

Figure 68. Extracted threshold voltage vs. back gate voltage

These effects are displayed in Figure 68, a plot of threshold voltage, extracted at a fixed drain current of 0.01uA/um, as a function of back gate bias. The data in the moderate back gate bias regime shows a somewhat linear character as the threshold voltage is truly modulated. In the high back gate bias regime, the curve appears slightly nonlinear due to the additional back channel current effect. Finally, in the low back gate bias regime, the curve also appears slightly nonlinear due to the degraded subthreshold slope effect. These electrical results indicated that many of the integrated process modifications were successful in improving the behavior of the IG-

FinFET structure. This electrical characterization represented the first demonstration of functional N-type IG-FinFET behavior [33].

6.5 Results of PFET2c1

After several fabrication experiments and integrated process changes, such as the change from PECVD Nitride to LPCVD Nitride for the gate etch and source/drain implant mask, two full PFET fabrication runs were completed. Electrical results were obtained for many devices from the PFET2c1 run. Several devices with gate lengths of 0.35um and 0.25um yielded electrical data, indicating that improvements in CMP processing and PC lithography had resulted in the successful fabrication of short channel FETs with acceptable PC-RX overlay tolerance. Devices from this run showed extremely high off-current resulting from misprocessing in a rapid thermal anneal (RTA) step as shown in Table 5. This misprocessing caused the source/drain junctions to diffuse substantially further under the gate electrode, and in fact, short together in a small portion of the effective width of the device. This area of shorting created a small current path. The MOS current from the remainder of the device is strong enough to overcome the shorting current, as seen in Figure 69.

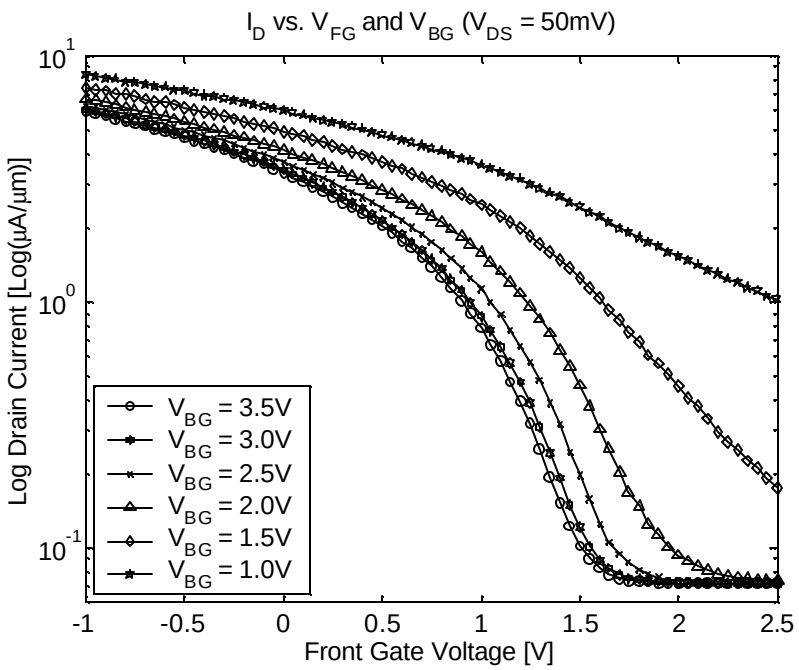

Figure 69. I_D vs. V_{FG}/V_{BG} at V_{DS} = -50mV (linear mode) for T_{FIN} = 50nm, L_{GATE} = 0.35um, Gamma = 0.25um device from PFET2c1 process run.

Device simulations were run to determine the percentage of effective width that would have to be shorted to produce device curves similar to those recorded in the characterization of PFET2c1 devices. Simulation results indicated that if 2% of the overall effective width were shorted, the remaining area would not produce enough MOS current to overcome the shorting current. If roughly 0.5% of the overall effective width were shorted, then the device curves would approximate those measured. This indicated that roughly 4-6nm of the overall effective width, translating to 2-3nm of the height of the fin was causing this short. This is probably again due to a non-uniformity in the source/drain dopant profiles. If substantially more dopant was implanted into the top of the fin, it is conceivable that the junction will diffuse faster at the top of the fin than on the sides of the fin, creating the short at the very top of the fin, with a functional device on the sidewalls.

6.6 Results of PFET2c2

After correcting the misprocessed RTA conditions from PFET2c1, another PFET run was completed. The results from this run were extremely positive. First, the yield from this run was extremely high. Many devices showed functionality and proper behavior. This enabled detailed trend analysis to see the effects of design dimensions such as fin thickness and extension length (Gamma). Second, the extracted results from these devices closely matched simulation results, indicating a successful integration. The results represented the first functional P-Type IG-FinFET results reported [34].

6.6.1 Double-Gate Mode Characterization

In order to first characterize the condition of the FET structure, double-gate mode characterization was conducted. In this experiment, all four terminals (Front Gate, Back Gate, Source and Drain) are probed independently, but the Front Gate and Back Gate probes are shorted and connected the single gate SMU from the HP 4156A. The intent of this experiment is to characterize the FinFET as if the gates had never been separated, to verify the character of the double-gate MOS structure. Typical I_D vs. V_{GS} curves were measured, and device data was extracted from the curves.

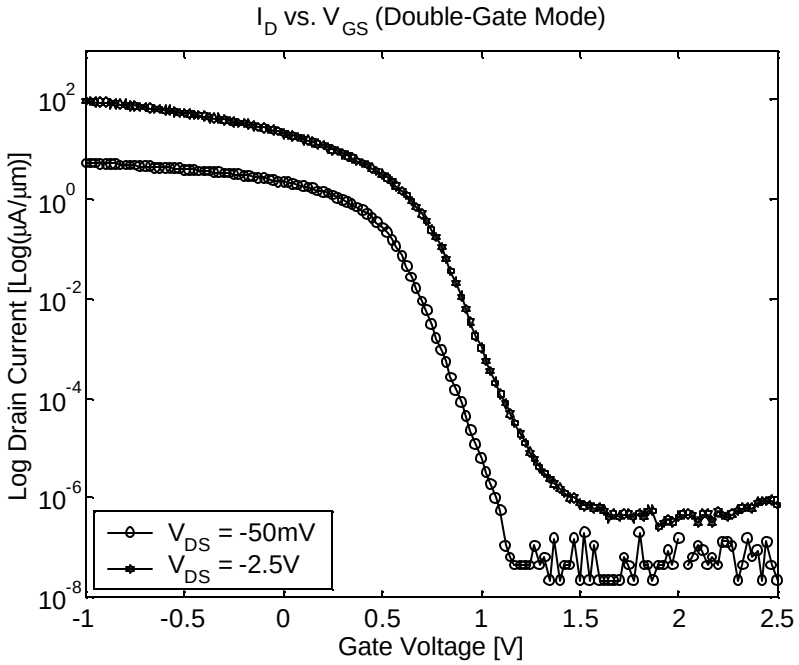

Figure 70. I_D vs. V_{GS} at V_{DS} = -50mV (linear) and -2.5V (saturation) for T_{FIN} = 50nm, L_{GATE} = 0. 5um, Gamma = 0.25um device from PFET2c2 process run.

The subthreshold slope for the device shown in Figure 70 is roughly 89mV/dec (linear) and 98mV/dec (saturation). The Drain Induced Barrier Lowering (DIBL) measured at I_D = 10nA/um is approximately 81mV/V. This DIBL measurement indicates that the effective channel length is significantly shorter than the drawn gate length of the device (L_{DRAWN} = 0.50um). Results of device simulations suggest that this amount of DIBL would be appropriate for this device with light body doping and no halo implants at a channel length near 0.20um. If the electrical channel length of this device were 0.20um, the length of the source/drain regions overlapping the gate edge would have to be 0.15um on each side. This hypothesis is supported by the fact that most of the 0.35um and all of the 0.25um drawn gate length devices from PFET2c2 run are shorted from source to drain, as they would have resulted in 0.05um (marginally shorted) and -0.05um (definitely shorted) effective gate length devices given a 0.15um overlap on each side.. This hypothesis is further supported by long

channel device data showing nearly no DIBL, as seen in Figure 71, indicating that the DIBL on the 0.5um drawn gate length device is caused by SCE and not an additional intrinsic device effect.

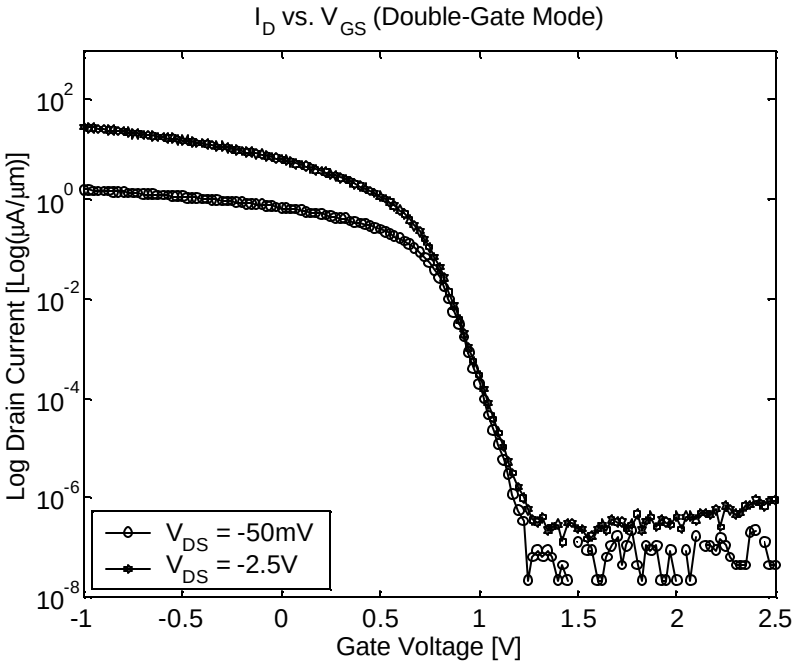

Figure 71. I_D vs. V_{GS} V_{DS} = -50mV (linear) and -2.5V (saturation) for T_{FIN} = 50nm, L_{GATE} = 2um, Gamma = 0.25um device from PFET2c2 process run.

Drain current was also measured as a function of drain voltage, for various gate voltages. This data, as seen in Figure 72, shows some slight effect of series resistance in the linear regime. This resistance is due to the thin extension region, and the lack of Silicide or a deep source/drain implanted region. However, due to the overlay accuracy in PC lithography, the extension region is kept short (Gamma = 0.25um), and therefore the device performance is not severely impacted. This data also shows that the contacts to the source and drain are ohmic, even without a heavily implanted source/drain or Silicide process.

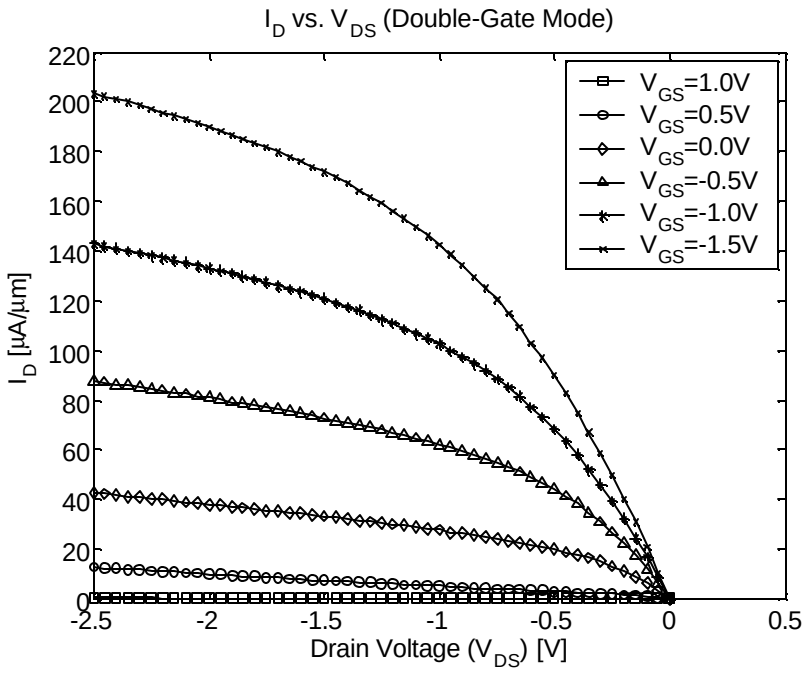

Figure 72. I_D vs. V_{DS} for T_{FIN} = 50nm, L_{GATE} = 0.5um, Gamma = 0.25um device from PFET2c2 process run.

The I_D vs. V_{DS} curves show a strong PFET drain current of 172uA/um at V_{GS} = V_{DS} = -1.5V and 203uA/um at V_{GS} = -1.5V and V_{DS} = -2.5V. Since the threshold voltage of these PFETs is positive (roughly 700mV positive), it is proper to examine the currents as a function of the gate overdrive (gate voltage applied beyond threshold conditions), and not as a function of the absolute gate voltage. In this respect, the V_{GS} = -1.5V condition is approximately 2.2V of overdrive, consistent with the reporting of PFET devices with negative threshold voltages on a 2.5V scale

. Examining data taken from several devices across the range of design parameters yielded additional insight into the performance of the devices on this run.

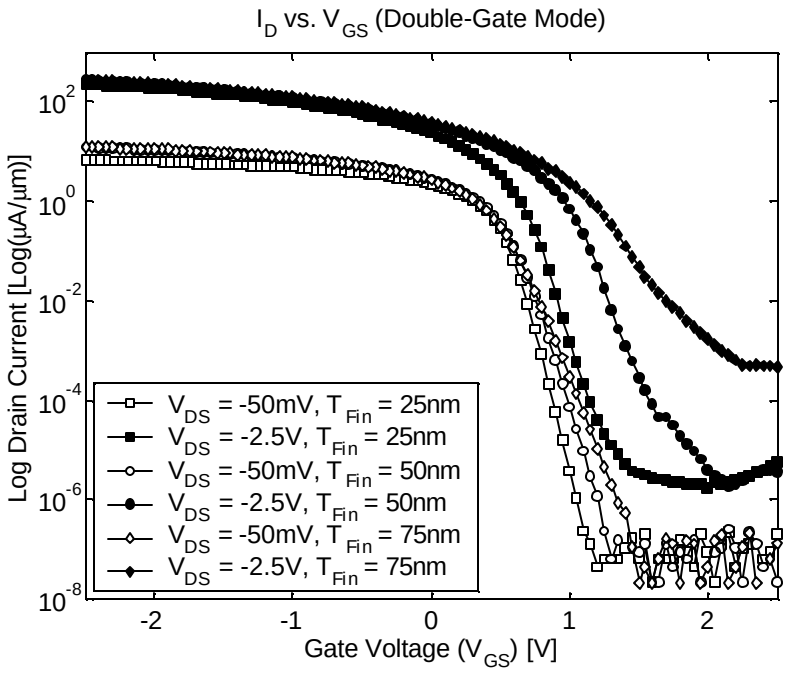

Figure 73. I_D vs. V_{GS} V_{DS} = -50mV (linear) and -2.5V (saturation) for multiple T_{FIN} dimensions, L_{GATE} = 0.5um, Gamma = 0.25um device from single die on PFET2c2 process run.

From the linear mode device curves shown in Figure 73, it is clear that the fin thickness has a strong effect on the subthreshold slope of the resulting device. The devices fabricated on thinner fins exhibit substantially better turn-on characteristics and subthreshold slope. This is expected due to the increase in gate to inversion layer coupling. The fin thickness also has a strong effect on the DIBL. The devices fabricated on thicker fins exhibit higher DIBL, and severely degrades saturation-condition turn-on characteristics. This too is expected, as the thicker fin affords more drain to channel depletion area, and therefore a stronger drain to inversion layer capacitance.

6.6.2 Double-Gate Mode Trend Analysis

More information can be displayed by extracting these parameters from a larger sample of devices and examining the data in a statistical fashion. Many of the devices from the PFET2c2 yielded functional device results.

Table 6. Design dimensions of functional devices measured across three die from PFET2c2 process run

# of Devs	T_{FIN}=25nm	T_{FIN}=50nm	T_{FIN}=75nm	***Total***
L_D=0.5um	**6**	**6**	**6**	**18**
L_D=2um	**6**	**6**	**5**	**17**
L_D=5um	**4**	**5**	**4**	**13**
Total	**16**	**17**	**15**	**48**

Trends are more easily displayed by examining the extracted device data, and plotting the mean values of all devices sampled as a function of designed device parameters.

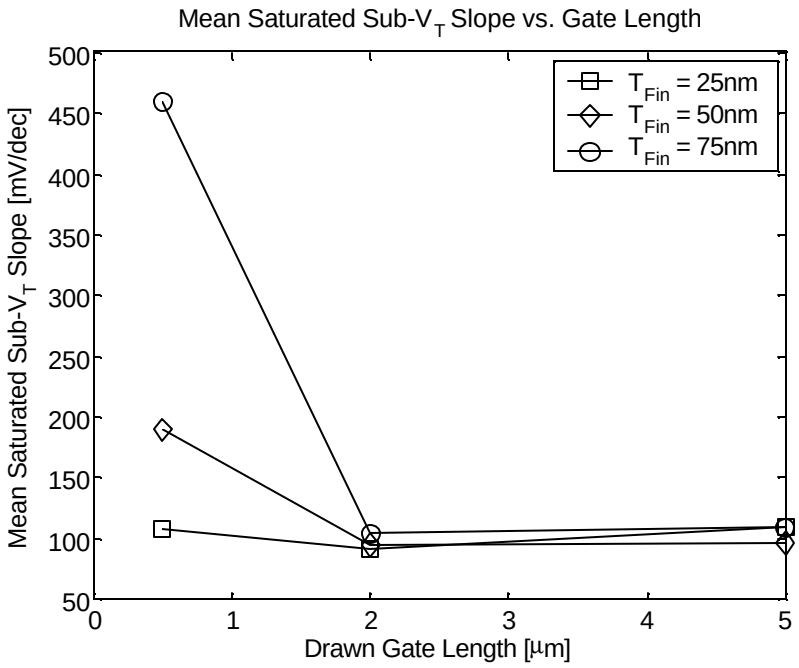

Figure 74. Mean saturated subthreshold slope vs. drawn gate length for devices fabricated in PFET2c2 process run.

This statistical data confirmed the earlier analysis on one die site. The subthreshold slope is severely degraded for the short channel devices fabricated on thinner fins, as seen in Figure 74. No degradation is visible on the 2um and 5um drawn gate length devices where the short channel effects are expected to be negligible. DIBL is also negatively impacted on devices fabricated on thicker fins, as seen in Figure 75. Again, no degradation is visible at the longer channel lengths.

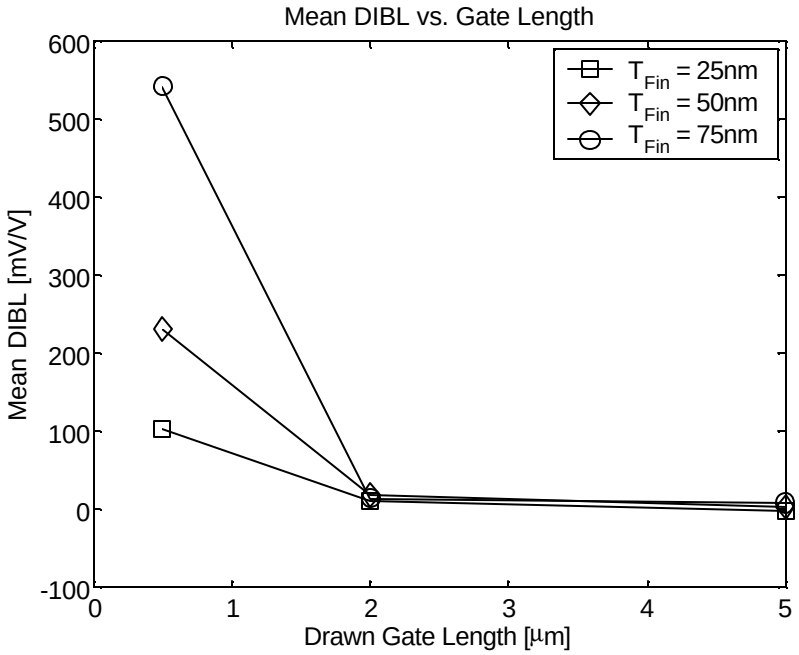

Figure 75. Mean Drain Induced Barrier Lowering (DIBL) vs. drawn gate length for devices fabricated in PFET2c2 process run.

The same effects responsible for degrading the subthreshold slope and DIBL on short channel devices on thicker fins can be measured as changes in the threshold voltage and off current. Since the threshold voltage is extracted at a fixed current density, an increase in DIBL is measured as an increase in threshold voltage. As seen in Figure 76, the threshold voltage roll-up is significantly worse for devices fabricated on thicker fins. The roll-up is almost negligible on the devices fabricated on 25nm fins, indicating the superior control over SCE delivered with a thin silicon body DGFET. As seen in Figure 77, the off-current density also increases significantly for short channel devices fabricated on thicker fins. This is simply another manifestation of the decrease in control over SCE in the thicker fin devices.

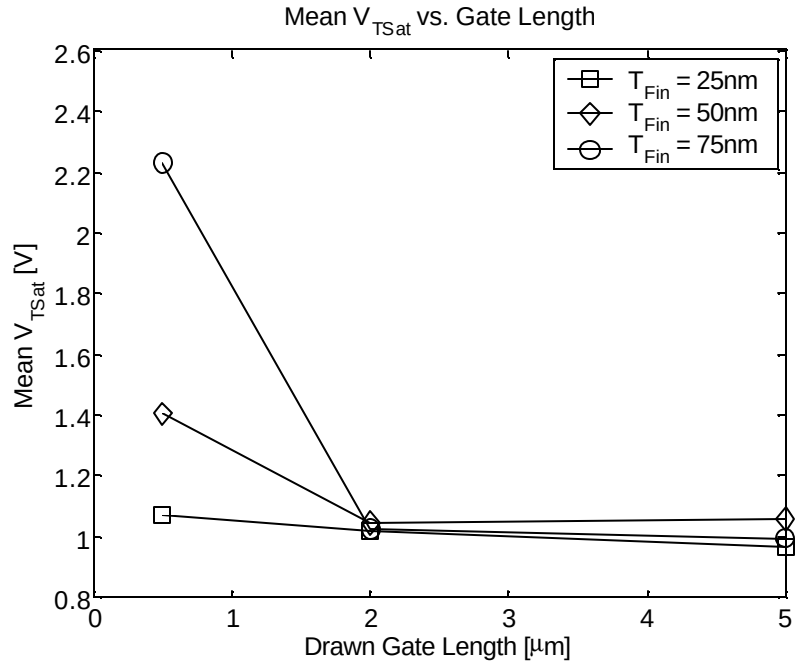

Figure 76. Mean saturated threshold voltage (V_{TSat}) vs. drawn gate length for devices fabricated in PFET2c2 process run.

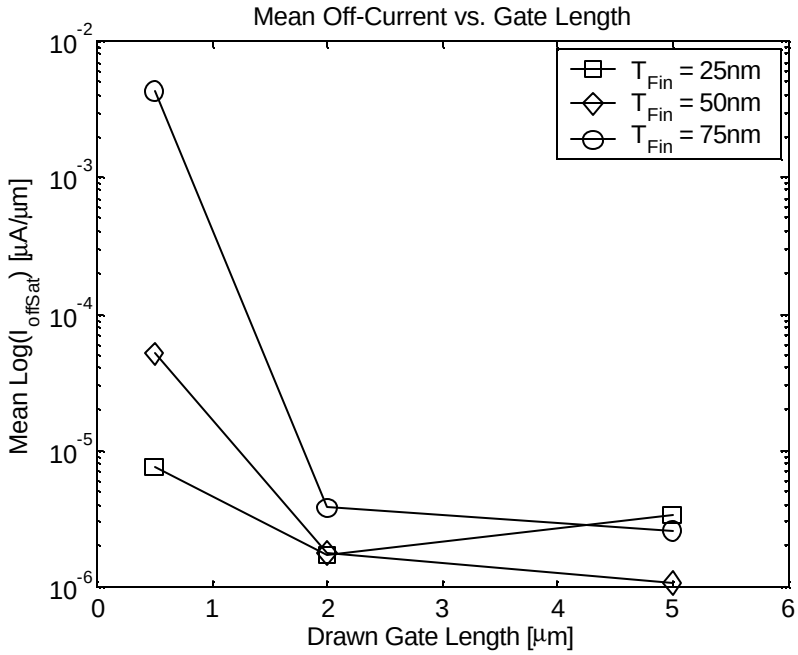

Figure 77. Mean saturated off-current density (I_{OFFSat}) vs. drawn gate length for devices fabricated in PFET2c2 process run.

All of the statistical data confirms simulation data and physical expectations of improved short channel control with thinner silicon fins. One unfortunate consequence of the thinner fin in this experiment is the increased source/drain resistance. Since there was no Silicide process used to decrease source/drain resistance, the resistance per unit length of the extension regions is simply a function of cross sectional area and doping density. As a result, the devices fabricated on thinner fins show improved short channel control, but lower saturated on-current density due to parasitic source/drain resistance, as seen in Figure 78.

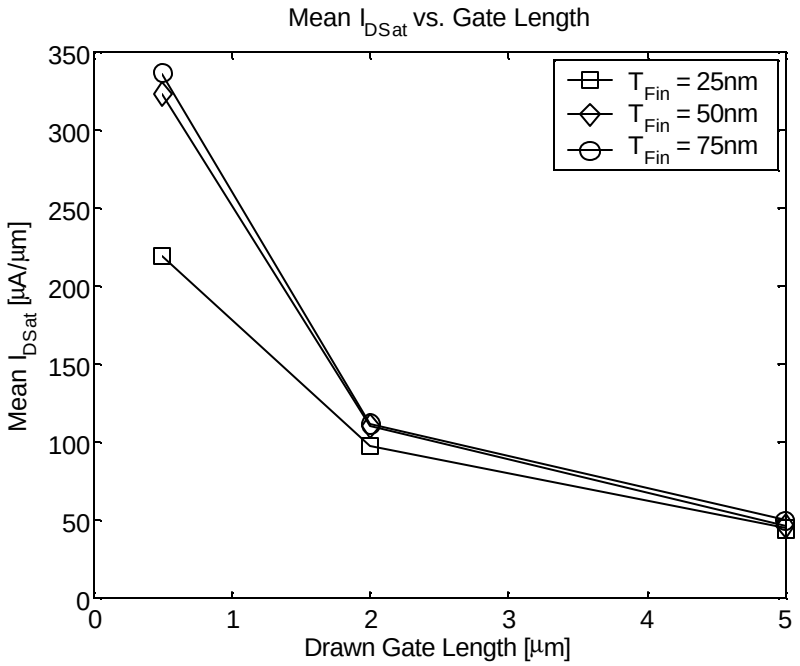

Figure 78. Mean saturated on-current density (I_{DSat}) vs. drawn gate length for devices fabricated in PFET2c2 process run.

Minimal degradation is seen between the devices fabricated on 75nm and 50nm drawn fin thicknesses. However, a substantial decrease is seen in the devices fabricated on 25nm drawn fin thicknesses. This may indicate that the source/drain dopant ions did not penetrate all the way through the fin thickness of the thicker fins,

and therefore the majority of the device current is being carried by a similar volume of doped silicon in the 50nm and 75nm fin thickness devices. If this were the case, then the parasitic resistance degradation would only occur to devices with fin thickness below that where the dopant would totally penetrate the fin.

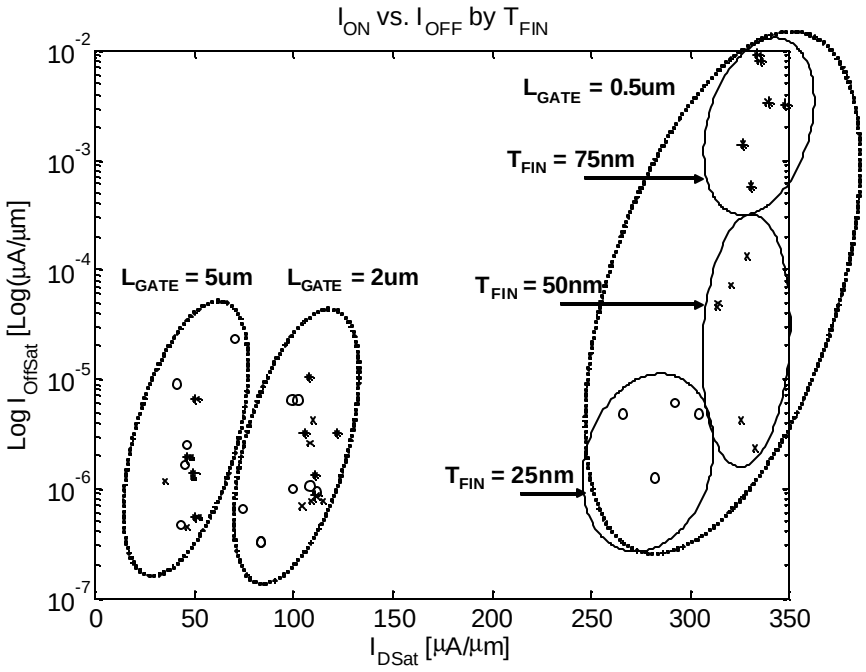

Figure 79. I_{ON} vs. I_{OFF} for all devices sampled in PFET2c2 run

Observation of the saturated on current as a function of the saturated off current, as seen in Figure 79, shows the typical trend as a function of gate length. However, an additional trend is observed as a function of fin thickness for the short channel devices. This is expected, and is in fact another manifestation of the poor short channel control of the thicker fin devices. The combination of these trends and relative density of data serve to support the observations that these IG-FinFETs, when biased in double-gate mode are performing as expected, and demonstrating the proper characteristics of double-gate CMOS devices.

6.6.3 Independent-Gate Mode Characterization

In order to verify complete gate separation, and independent-gate behavior of the PFETs fabricated on the PFET2c2 run, I_D vs. V_{FG}/V_{BG} measurements were taken for both linear and saturation conditions. As expected, based on the successful double-gate mode characterization, the independent-gate measurements showed excellent device behavior.

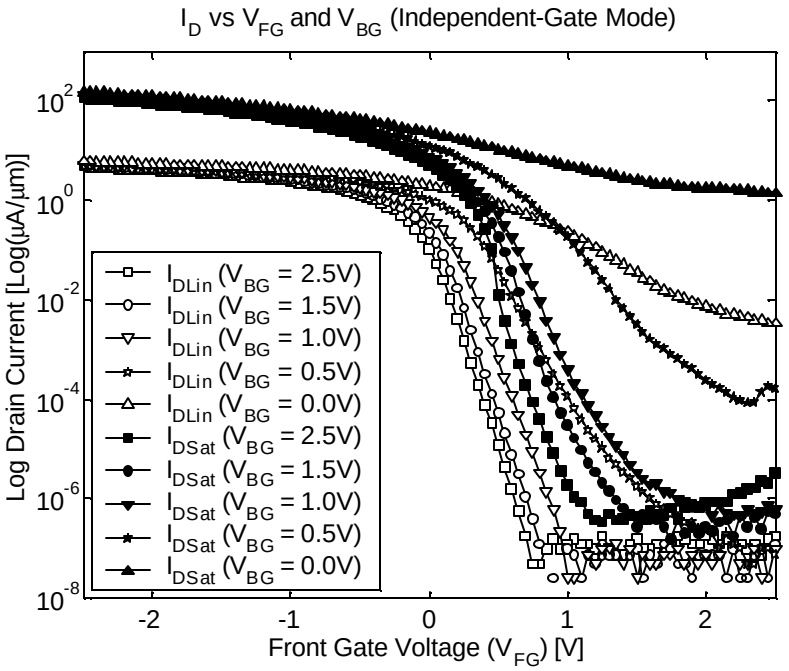

Figure 80. I_D vs. V_{FG}/V_{BG} at V_{DS} = -50mV (linear) and -2.5V (saturation) for T_{FIN} = 50nm, L_{GATE} = 0. 5um, Gamma = 0.25um device from PFET2c2 process run.

As predicted by simulation, the back gate bias acts to adjust the threshold voltage of the front gate, until, at high enough back gate bias, the back channel enters inversion, and the back channel current prevents the front gate from turning the aggregate device off. The threshold modulation is more easily visible in a linear scale plot, as seen in Figure 81.

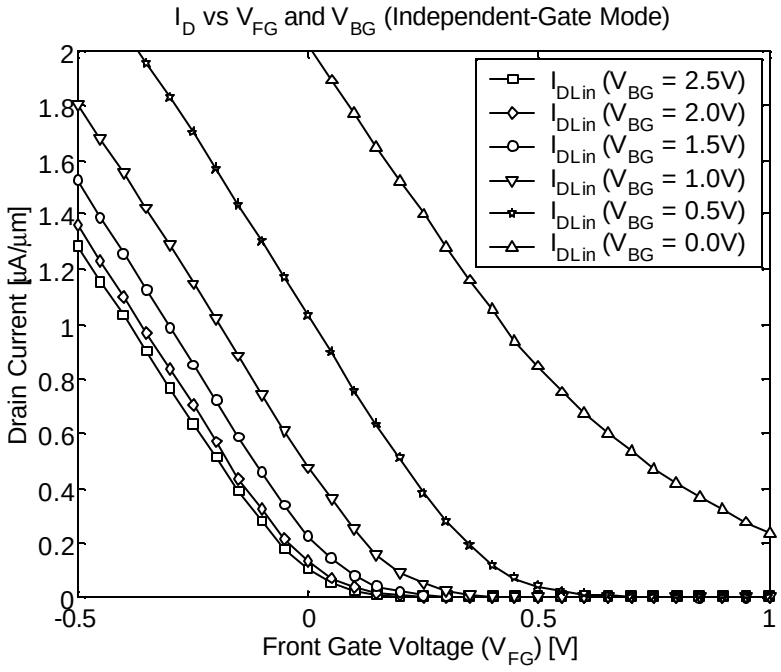

Figure 81. Linear scale I_D vs. V_{FG}/V_{BG} at V_{DS} = -50mV (linear) for T_{FIN} = 50nm, L_{GATE} = 0. 5um, Gamma = 0.25um device from PFET2c2 process run focusing on threshold region.

The modulation of the front gate threshold voltage by way of back gate bias shows that the gate separation was successful, and that central double-gate structure is functioning as expected. Negligible measured gate current also verifies that the gate separation is complete, with no residual gate polysilicon remaining on the top of the fin hardmask. By extracting the threshold voltage from the device curves using a linear intercept extraction method, a measure of threshold tuning can be determined. Again, as seen in Figure 82, the threshold voltage is strongly coupled to the back gate bias voltage. The DIBL resulting from SCE, as discussed in Section 6.6.1, is evident again in the separation between these curves. However, the additional functionality of threshold voltage tuning is clearly displayed.

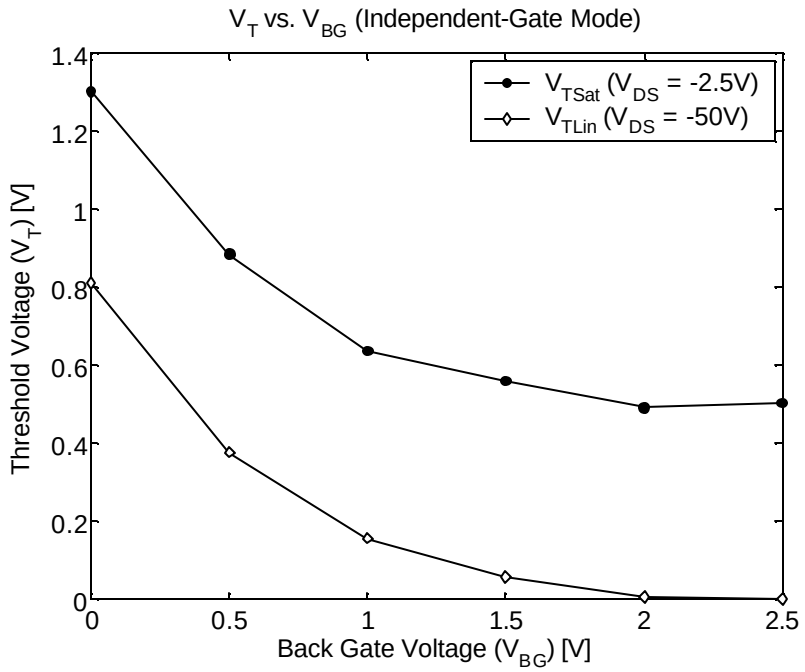

Figure 82. Threshold Voltage vs. Back-Gate Voltage for T_{FIN} = 50nm, L_{GATE} = 0.5um, Gamma = 0.25um device from PFET2c2 process run.

A figure of merit can be extracted by measuring the dependence of the threshold voltage on the back gate voltage through the linear regime of the plot. An example of the extraction of this type of metric is shown in Figure 83. To eliminate the effects of DIBL on the thicker fin devices, linear threshold voltage is used for this extraction. The delta is calculated as the difference between the linear threshold voltage at two different back gate biases (in this case V_{BG1} = 2.0V, V_{BG2} = 1.5V) divided by the difference in back gate biases to arrive at a unit of V/V. As expected, the devices fabricated on thinner fins show a larger amount of threshold voltage tuning. The separation is extremely large on the long channel devices where the large gate areas provide for large capacitances relative to the junction capacitances for controlling the threshold voltage. The separation of the means as a function of fin thickness is smaller for the short channel devices as the junction profile plays a larger relative role in controlling the threshold voltage.

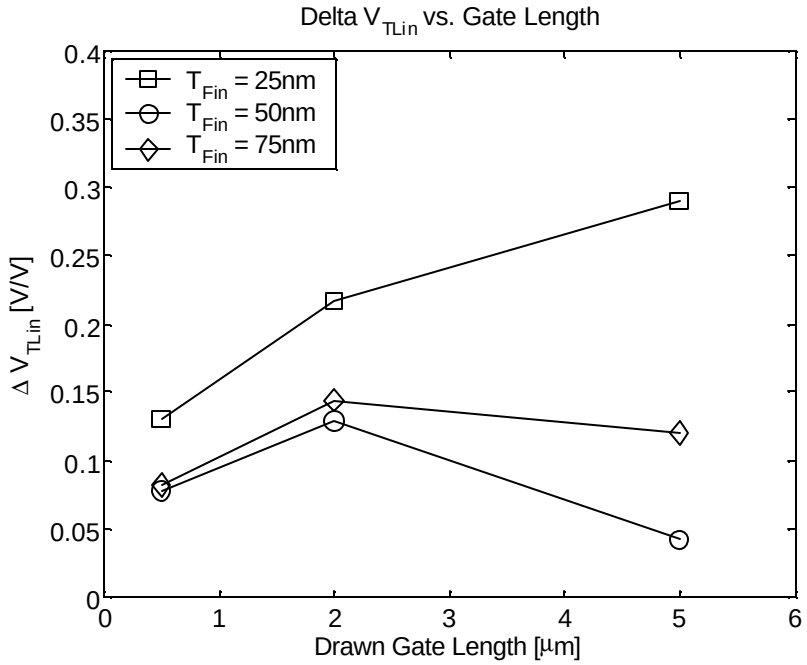

Figure 83. Change in V_{TLIN} as a function of device dimensions.

6.7 Conclusions

Device results are demonstrated showing the functional behavior of independent-gate FinFETs. Results from early process runs indicated substantial structural integration issues. These indications were verified with physical failure analysis. Data from later runs showed the effects of integrated process changes on the resulting devices. Finally, the first successful integration of two fully self-aligned independent gates on a FinFET is demonstrated, illustrating excellent double-gate and independent-gate behavior. The results from the final run show excellent subthreshold slope on long and short channel devices. Moderate DIBL on the short channel devices is believed to be due to excessive overlap, light body doping, and the absence of halo implants. High yield on the final run enabled trend analysis to study the effect of device dimensions on device performance. These trends show good agreement with simulations. The fin thickness is the most critical dimension for high performance devices in both double-gate and independent-gate mode. Thinner fins enable

improved short channel control over subthreshold slope, DIBL and absolute threshold voltage control. The amount of threshold tuning available is also increased in devices fabricated on thinner fins. These results show the possibility of integrating fully-self aligned independent double-gate devices in a manufacturable fin-based CMOS technology.

Chapter 7: Conclusions and Suggestions for Further Research

7.1 Summary

The Independent-Gate FinFET has been introduced as a novel device structure that combines several innovative aspects of the FinFET and planar double-gate FETs. The IG-FinFET addresses the concerns of scaled CMOS at extremely short channel lengths, by offering the short channel control of the double-gate architecture. The IG-FinFET allows for the unique behavioral characteristics of an independent-gate, four-terminal FET. Finally, this device also allows for conventional CMOS manufacturing techniques to be used by leveraging many of the FinFET integration concepts. By introducing relatively few deviations from a standard FinFET fabrication process, the IG-FinFET integration offers the capability of combining three-terminal FinFET devices with four-terminal IG-FinFET devices in one powerful technology for SoC or analog/RF application, to name only a few. This device has been examined by device modeling, circuit simulation, testsite design, fabrication and electrical characterization.

7.1.1 IG-FinFET Device and Circuit Design

The Independent-Gate FinFET has been examined using 2-D device simulation, both on the device stack and a full transistor structure. The effects of gate-misalignment have been explored to better understand the desire for full self-alignment in double gate structures. Circuit designs have been examined using a

quasi-static device model and a SPICE based circuit simulator. These circuit simulations show the potentially powerful aspects of the independent gate architecture. An examination of the physical layout aspects of both nominal FinFETs and IG-FinFETs has been done to understand the layout efficiency of these devices. The physical design of these circuits highlights a potential drawback to the IG-FinFET, in the severe layout penalty of an IG-FinFET with large effective device width. However, the nominal FinFET is shown to have the capability of being more layout efficient even than planar devices. The combination of these findings leads to the motivation for a fully self-aligned independent-double-gate FET architecture that can be easily integrated with a more layout efficient double-gate device. The IG-FinFET satisfies all of these requirements.

7.1.2 Fabrication of IG-FinFETs

A test vehicle has been designed for the process development, integration, and hardware validation of the Independent-Gate FinFET concept. This test vehicle contains many of the structures necessary to test and characterize the many modules of the fabrication process, as well as many devices and structures to validate the performance and behavior of this novel device. A fabrication process has been developed, beginning from an initial conceptual scheme, through experimentation and subsequent process development, to a successful structural integration. Detailed characterization has been performed to study many of the modules of the fabrication process, and significant development has been conducted on the process flow. The Chemical Mechanical Polish of the gate electrode has been carefully examined, and reconfigured to improve device yield and uniformity. The implant masking by the gate electrode hardmask has also been investigated, and process changes have been implemented to improve the quality of this implant mask. This fabrication flow has

been validated for the integration of NFET and PFET devices, and provides a platform for the study of several aspects of the device architecture. This complete fabrication process has been implemented and documented to provide a FinFET and IG-FinFET technology platform for the Cornell Nanofabrication Facility, and ongoing research in the field of double-gate devices.

7.1.3 Electrical Characterization of IG-FinFETs

Electrical characterization has been performed at several stages in the integrated process development. Characterization of early fabricated devices and structures focused on failure analysis. Electrical results from these devices corroborated evidence from physical failure analysis to direct changes to the fabrication process. Continuing characterization on later hardware showed evidence of the success of these process changes. Electrical characterization of the final fabricated hardware showed excellent agreement with simulation and confirmation of the success of the integration scheme. The many process changes that were discovered and implemented in earlier runs were validated on the final process run. Devices from the final set of experiments showed excellent double-gate behavior, with steep subthreshold slopes and low off currents. Also, these devices showed experimental evidence of the device concepts discussed in the device modeling and simulation chapter. The devices fabricated on thinner fins exhibited lower subthreshold slope, lower DIBL, but also lower on current, ostensibly due to the increased parasitic source/drain resistance of the thinner extension region. When tested in Independent-Gate mode, these same devices showed excellent agreement with IG-Mode simulation. This provided experimental evidence of the success of the gate separation and isolation technique, as well as the source/drain implant masking by use of the gate electrode etch mask. The IG-Mode extracted data also corroborated

device theories. Devices fabricated on thinner fins showed improved short channel control, and improved threshold voltage control, as expected.

7.2 Suggestions for Further Research

7.2.1 Fabrication Process Optimization

This work was aimed at a first demonstration of this novel device architecture. In order to produce a more repeatable, large scale technology, several unit processes must be optimized. The control of the fin thickness and line edge roughness must be examined. Since these parameters affect threshold voltage and carrier mobility, their control is paramount to the large scale integration of this device. There is already ongoing research to optimize the fin surface using plane-selective wet etching [35]. While this process may improve the surface of fin sidewalls designed on or near optimal planes, it severely limits the designs of FinFET circuits. Control of these tolerances may be achieved using Sidewall Image Transfer, without limiting the design capabilities of the device.

This work attempted to choose source/drain implants that would yield functional devices at several channel lengths. More development is required to determine optimal implant conditions for this device given its unique three dimensional structure and the fact that different materials provide the implant mask on different parts of the device. Three-dimensional process simulation will be required to fully understand the effects of the structure on final junction profiles. Also, a large amount of hardware verification will be needed to verify these simulations as the thin-vertical body may have additional effects on the implant that may not be captured by the process models. Finally, a true source/drain junction will almost certainly require an LDD architecture, perhaps with halo implants as well. This architecture is not trivial given the three dimensional nature of this device. The combination of these

implants, and subsequent effect on the final junction profile after anneals will be difficult to ensure. Advanced simulation and hardware verification will be required for this work.

7.2.2 CMOS Integration

Many of the structural elements required for full CMOS IG-FinFET integration are validated in this work. A few simple process changes will be required for CMOS integration, including block level lithography for dual gate pre-doping implants and complementary source/drain implants. However, CMOS integration will also drive a few more complex integration challenges. CMOS integration will almost certainly require a Salicide/Silicide technology. This creates requirements on the source/drain doping profiles. The source/drain must be heavily doped to produce a reliable ohmic contact with the Salicide/Silicide material. This would require some type of spacer process be implemtned. Due to the three dimensional nature of the structure, this is extremely delicate. A spacer must be formed in three locations of the device, (top of gate hardmask to top of fin, top of extrinsic gate to BOX, and in the corner of the gate and fin junction), while removing the spacer on the fin extension region. Since the three spacer structures required all have different heights, this process will require detailed development and characterization.

Finally, for CMOS and circuit integration, the design technique of flaring out the source/drain region to land a contact will be too layout inefficient. An architecture supporting direct metal to fin contacts will be required. This is also a difficult process to develop, as many conventional metal processes would damage the fin extension, causing yield fallout.

7.2.3 Circuit Characterization

Following the work of structural integration, process optimization and CMOS integration, circuit verification and characterization will be required. This work examined simple circuit designs aimed at understanding the device parameters of interest, and the implication of integrating this device with nominal FinFETs. Large scale circuit implementation for SoC and analog/RF applications will require more advanced detailed characterization of fabricated circuits. The IG-FinFET inverters and ring oscillators designed on the testsite vehicle for this work would provide a strong initial set of structures for characterization. But large devices and circuits for further characterization of the device and its behavior in circuits will have to be designed and laid out appropriately.

Appendix A: Testsite Design Images

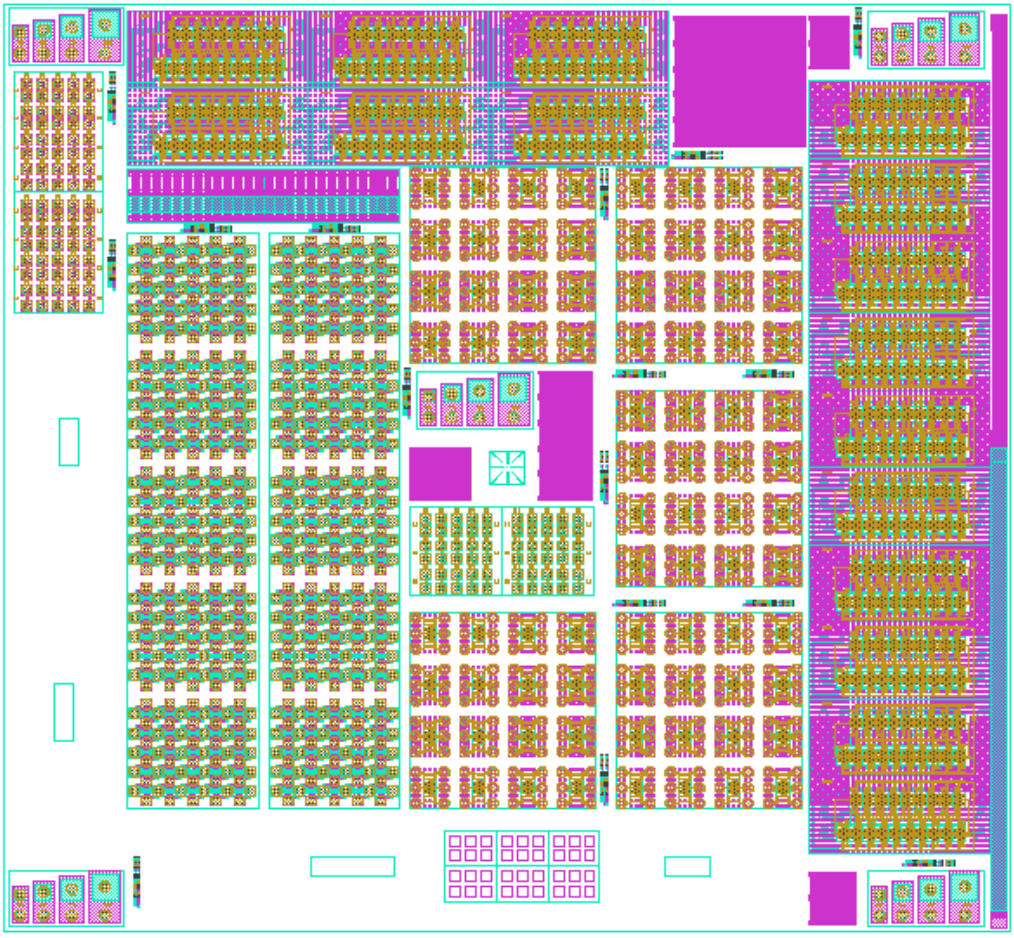

Figure 84. Full Testsite Layout

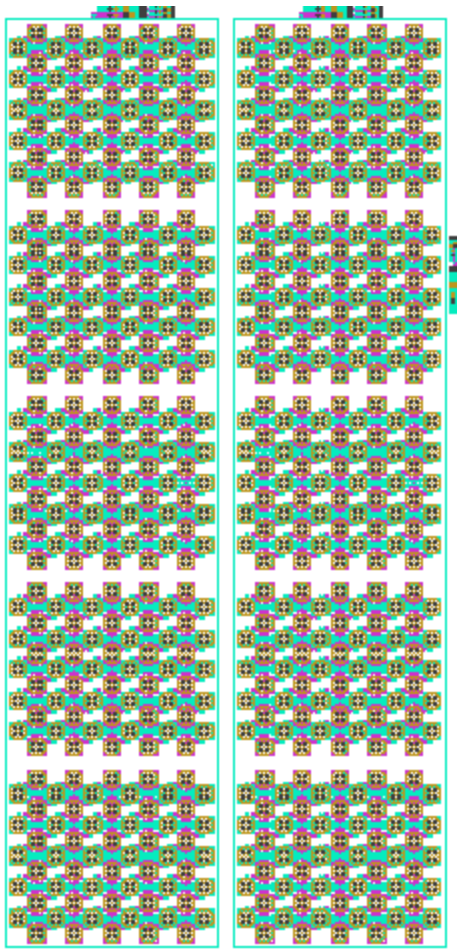

Figure 85. Device Arrays (NFETs on left, PFETs on right)

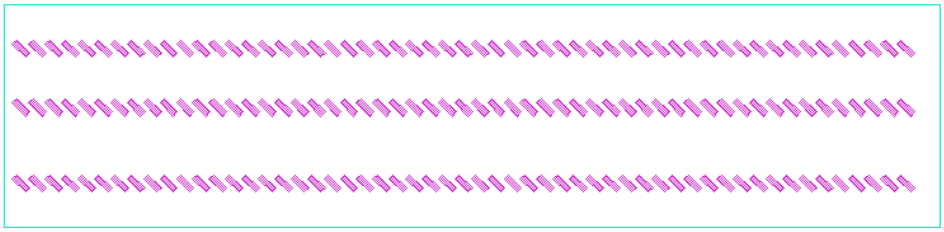

Figure 86. Structures included on RX level to permit DUV alignment to e-beam written features

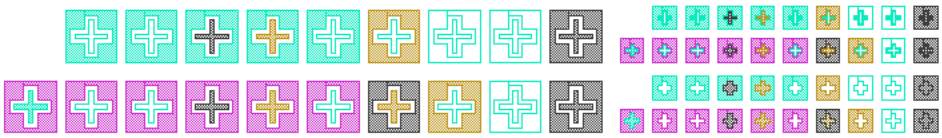

Figure 87. Structures included on many levels to measure level-to-level overlay

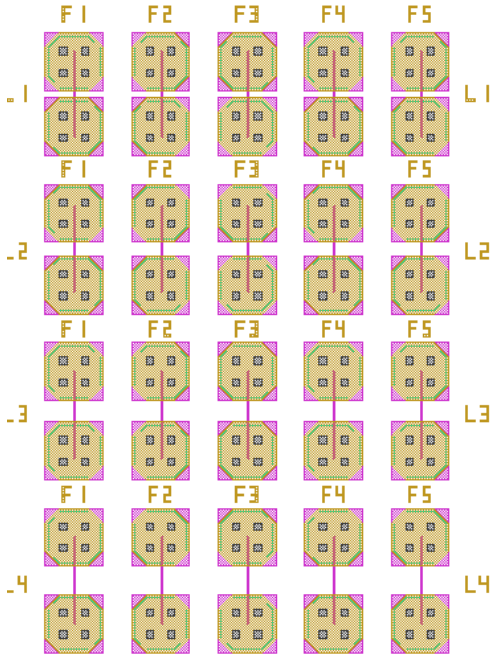

Figure 88. Fin resistance measurement macro design

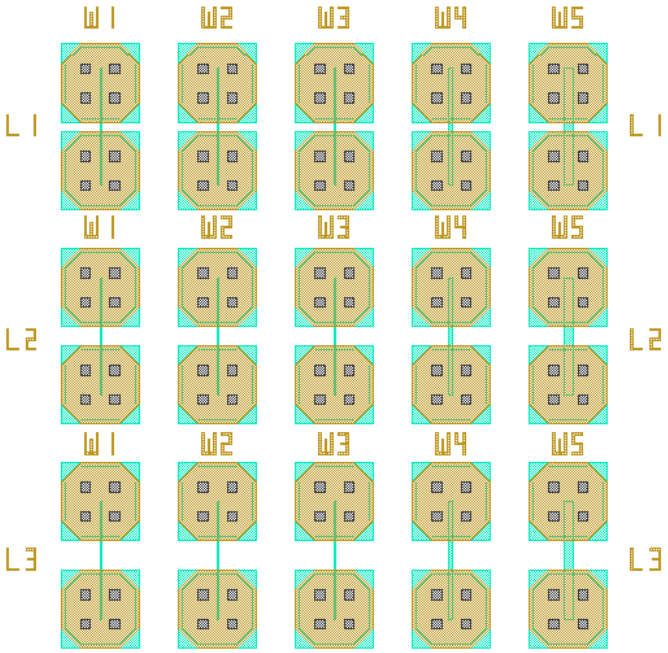

Figure 89. Polysilicon gate resistance measurement macro design

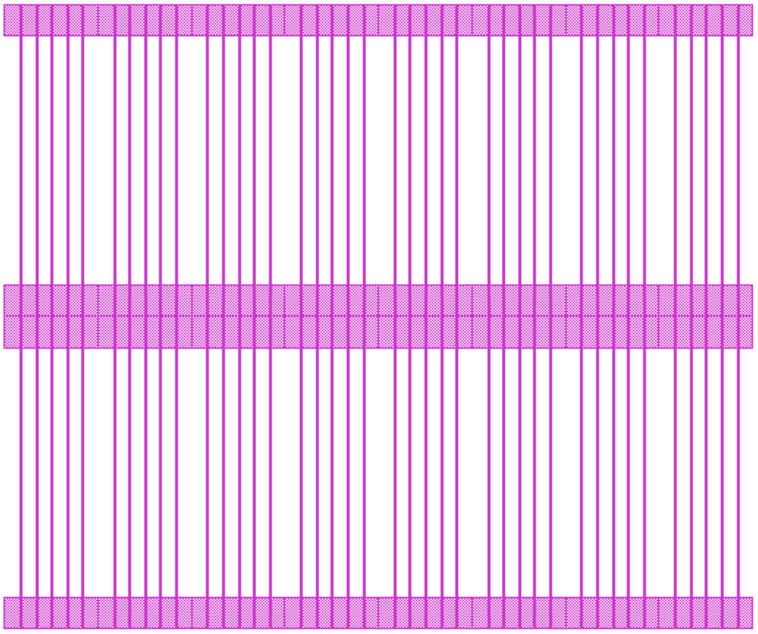

Figure 90. Profilometric measurement macro design

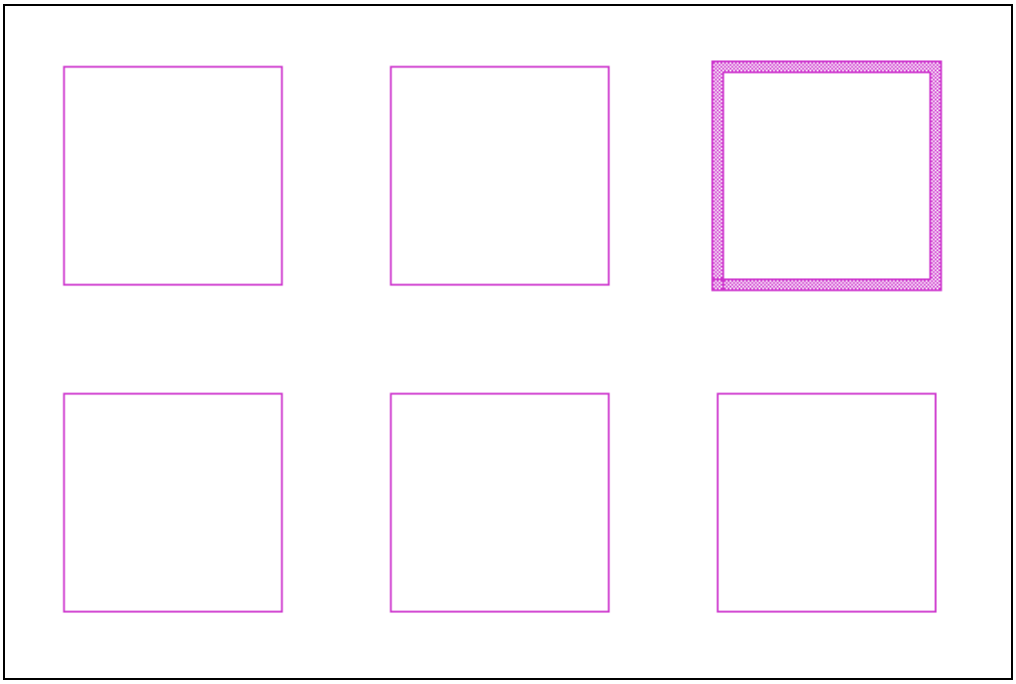

Figure 91. CMP completion measurement macro design

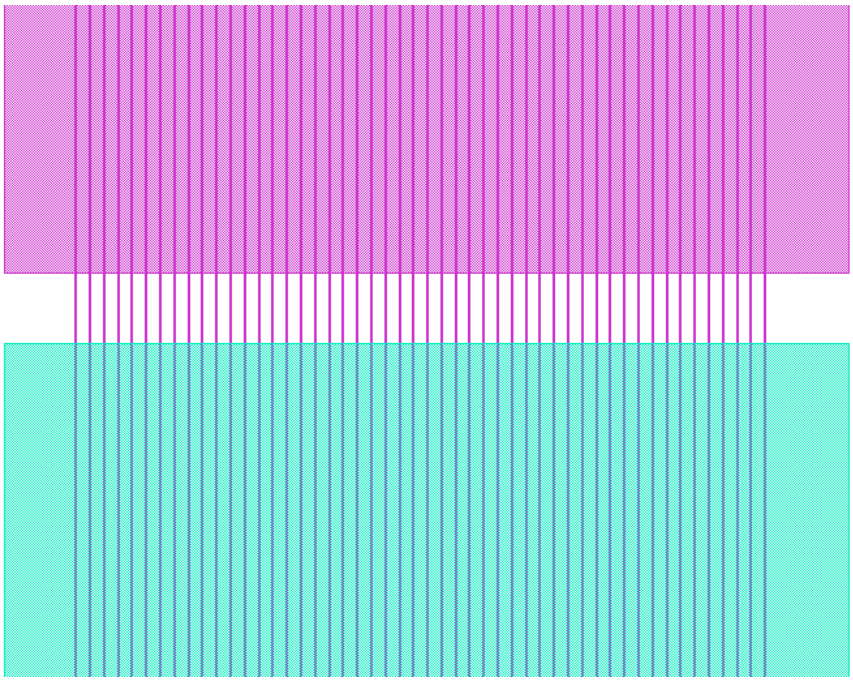

Figure 92. SEM Cross-Section measurement macro design

Appendix B: Electron Beam Lithography Job File

The following is the text of the electron beam lithography job-file. This file governs the behavior of the Leica VB6. It is set up to run two different FRE files (design data) with different doses, stepped out over an array. The array size, stepping distance, dose and design files can be edited to write different data onto the wafer. This file was generated from an existing file to do similar stepping [36]. Significant changes were required for this functionality and for the proper writing of this work. Minor formatting changes have been applied for the sake of this publication. These changes have no effect on the jobfile commands, simply on the comments.

```
$! THIS JOB FILE WRITES TWO DIFFERENT FRE FILES
$! WITH DIFFERENT CURRENTS
$! IT IS SPECIFIC TO DAVID FRIED'S FIN_TESTSITE FILES/CURRENTS
$! YOU MUST LOADFINE, FOC, JC, BEFORE RUNNING THIS JOB FILE
$!=======================================
$! DEFINITIONS
$!=======================================
$ WS :== WRITE SYS$OUTPUT
$! WRITE TO TERMINAL WINDOW
$!
$!=======================================
$! LOGGING
$!=======================================
$ WS ""
$ WS "**************"
$ WS "* START TIME *"
$ WS "**************"
$ WS ""
$ SHOW TIME
$ WS ""
```

```
$!
$!=======================================
$! CALIBRATION
$!=======================================
$ DWMO ABS /LOAD                        ! IN ABSOLUTE MODE
$ SFAB                                  ! FAB MODE, BEAM OFF
$ QSET EROSION NOERODE                  ! LEAVE SHAPES IN EXEL FORMAT
$ QSET CORR ON/ALL                      ! TURN ON ALL CORRECTIONS
$ QSET HEIGHT/REALTIME                  ! SET REAL TIME HGT CORR.
$ QSET SORT NORMAL                      ! NORMAL PATTERN SORTING
$!
$!=======================================
$! MAP TO WAFER CENTER
$!=======================================
$ MVSP CENTRE                           ! MOVE TO CENTRE OF WAFER
$ SSPO ZERO 0 0                         ! STORE 0,0 AS "ZERO"
$ DWCO REL /EXP=(ZERO) /OBS=(CENTRE)    ! REMAP ORIGIN TO CENTER
$                                       !IN REL MODE
$ DWMO REL /LOAD                        ! SWITCH TO REL MODE
$ MVPO 0 0                              ! MOVE TO 0 0 IN REL MODE
$ QDISPLAY HEIGHT/TAB=3                 ! TAKE A HEIGHT READING
$                                  ! /TAB=3 IS FOR SILICON: SEE MANUAL.
$!=======================================
$! DEFINE STEP PATTERN PARAMETERS
$!=======================================
$ NUMROWS = 6                      ! NUMBER OF ROWS
$ NUMCOLS = 6                      ! NUMBER OF COLUMNS
$ STEPX  := 5                      ! X PERIODICITY
$ STEPY  := 5                      ! Y PERIODICITY
$ STARTX := -12.5                  ! INITIAL X OFFSET
$ STARTY := -12.5                  ! INITIAL Y OFFSET
$ WS ""
$ WS
"=========================================================="
$ WS "THIS PATTERN WILL WRITE A "NUMROWS' X "NUMCOLS' MATRIX"
$ WS "WITH X PERIODICITY OF "STEPX' AND Y PERIODICITY OF "STEPY'"
$ WS "STARTING AT X = "STARTX' Y = "STARTY'"
$ WS
"=========================================================="
$ WS ""
$!
$!=======================================
$! SET DOSES FOR FIRST FILE
$!=======================================
$ PATRN :== [VB.USERS.FRIED]FIN_TESTSITE_41.FRE
```

```
$ QSET  VRU 2                       ! SETS THE VIRTUAL RESOLUTION UNIT
$ DOSE0      := 105                 ! INITIAL DOSE
$ DOSESTEP0  := 1                   ! DOSE STEP
$ OPERATOR   :=                     ! DOSE STEP METHOD
$ GOSUB EXPOSE_PATCH                ! RUNS THE STEP EXPOSURE ROUTINE
$!
$!=======================================
$! RECALIBRATE FOR THE SECOND FILE
$!=======================================
$ LOADCOARSE                        ! LOADS THE HIGHEST CURRENT IN DB
$ DWMO ABS /LOAD                    ! PUTS IT BACK INTO ABSOLUTE MODE
$ MVSP FM                           ! MOVES BACK TO THE FOCUS MARK
$ LOC FM FM /POSM                   ! JUST IN CASE IT CAN'T FIND IT
$ FOC                               ! FOCUSES THE BEAM
$ JC                                ! RUNS JOBCAL
$ DWCO REL /EXP=(ZERO) /OBS=(CENTRE)     ! REDEFINES THE OLD
$                                   !REL SYSTEM
$ DWMO REL /LOAD                    ! LOADS THE REL SYSTEM
$ MVPO 0 0                          ! MOVES TO 0 0 REL
$ QDISPLAY HEIGHT/TAB=3             ! TAKE A HEIGHT READING
$!                                  ! TAB=3 IS FOR SILICON: SEE MANUAL.
$!=======================================
$! SET DOSES FOR THE SECOND FILE
$!=======================================
$ PATRN :== [VB.USERS.FRIED]FIN_TESTSITE_43.FRE
$ QSET VRU 8                        ! SET THE VRU FOR HIGHER CURRENT
$ DOSE0     := 17                   ! INITIAL DOSE
$ DOSESTEP0 := 0.2                  ! DOSE INCREMENT
$ OPERATOR  := +                    ! INCREMENT METHOD
$ GOSUB EXPOSE_PATCH
$!=======================================
$! ALL DONE WRITING
$!=======================================
$ QSET EROSION NORMAL               ! RETURN THE EROSION SETTING
$ WS ""
$ WS "**************"
$ WS "* END TIME *"
$ WS "**************"
$ WS ""
$ SHOW TIME
$ WS ""
$!
$ EXIT      !!! END OF MAIN PROGRAM !!!
$!=======================================
$! BEGINNING OF EXPOSE_PATCH SUB-ROUTINE
```

```
$!=======================================
$ EXPOSE_PATCH:                 ! START OF DO LOOP
$ XX := 'STARTX'                ! DEFINES INITIAL X POSITION
$ YY := 'STARTY'                ! DEFINES INITIAL Y POSITION
$ MVPO /CORNER 'XX' 'YY'        ! GOES TO INITIAL X Y
$ SPAT 'PATRN'                  ! LOADS THE PATTERN
$ ROW = 0                       ! INIT. THE INCREMENT VARIABLE
$ FORROW: ! == 1 TO NUMROWS DO
$                       ! BEGINNING OF OUTER (ROW,Y) LOOP
$       ROW = ROW + 1
$       IF ROW .GT. NUMROWS THEN GOTO ENDFORROW
$       XX := 'STARTX'
$       COL = 0
$       FORCOL: ! == 1 TO NUMCOLS DO
$                       ! BEGINNING OF INNER (COL,X) LOOP
$               COL = COL + 1               ! THE REST IS SELF-EXPLANATORY
$               IF COL .GT. NUMCOLS THEN GOTO ENDFORCOL
$               WS ""
$               WS "======================================="
$               WS "DIE INDEX = ROW "ROW' COLUMN "COL'"
$               WS "MOVING STAGE TO X= "XX' Y= "YY'"
$               MVPO /CORNER 'XX' 'YY'
$               WS "SETTING DOSE TO "DOSE0'"
$               SHOW TIME
$               SDSE 0-31 'DOSE0' /ABS
$               ACLK /DOSE
$               WS "EXPOSING PATTERN= "PATRN'"
$               WS "======================================="
$               WS ""
$               EPAT
$!
$               WS ""
$               WS "======================================="
$               WS "CALCULATING NEXT X POSITION"
$               WS "-------------------------"
$               WS "PREVIOUS X= "XX'"
$               WS "X PERIODICITY= "STEPX'"
$               CL 'XX' 'STEPX' + XX
$               WS "NEXT X= "XX'"
$               WS "======================================="
$               WS ""
$!
$               WS ""
$               WS "======================================="
$               WS "CALCULATING NEXT DOSE"
```

```
$               WS "---------------------"
$               WS "PREVIOUS DOSE= "DOSE0'"
$               WS "DOSE INCREMENT= "DOSESTEP0'"
$               WS "DOSE INCREMENT METHOD= "OPERATOR'"
$               CL 'DOSE0' 'DOSESTEP0' 'OPERATOR' DOSE0
$               WS "NEXT DOSE= "DOSE0'"
$               WS "========================================="
$               WS ""
$!
$       GOTO FORCOL
$       ENDFORCOL:
$!
$       WS ""
$       WS "========================================="
$       WS "CALCULATING NEXT Y POSITION"
$       WS "PREVIOUS Y= "YY'"
$       WS "Y PERIODICITY= "STEPY'"
$       CL 'YY' 'STEPY' + YY
$       WS "NEW Y= "YY'"
$       WS ""
$!
$ GOTO FORROW
$ ENDFORROW:
$ RETURN
$!=======================================
$! END OF EXPOSE_PATCH SUB-ROUTINE
$!=======================================
```

Appendix C: Final Process Flow Listing

The following represents a comprehensive listing of the process steps required to fabricate the final uni-polar (either NFET-only or PFET-only) IG-FinFET experiment as discussed in Chapters 4-6. This process flow listing is optimized for fabrication in the Cornell Nanofabrication Facility, with the intent of potential transfer to other facilities.

Gate	Proc Type	Material or Mask	Tool	Process
RX Gate				
RX	Start	SOI	SOITEC	Starting SOI Wafer
RX	Clean	RCA	MOS	MOS Clean w/ HF Dip
RX	Growth	SiO2	Thermco	Wet Thermal Oxide Growth (Hardmask) (Proc 50, 1000C, 35min)
RX	Bake	H2O	HotPlate	Dehydrate Bake (170C, 5min)
RX	Clean	Acetone/IPA	EBSpin	Solvent Clean
RX	Spin	NEB-31	EBSpin	Spin E-Beam Resist (NEB-31, 4000rpm, 60sec)
RX	Bake	NEB-31	CEE	Pre-Bake (110C, 2min)
RX	Expose	CMPFR43	HTG	RX DUV CMP Balancing Exposure (90sec)
RX	Expose	FINRXMASTER	VB6	RX E-Beam Write (FINRXMASTER.COM)
RX	Bake	NEB-31	CEE	Post-Bake (95C, 4min)
RX	Develop	MF-321	Wet	Develop (MF-321, 35sec)
RX	Rinse	DI	Wet	Rinse in DI Water
RX	Dry	N2	N2 Gun	Blow Dry
RX	Etch	SiO2	AMAT	RX Thermal Oxide Cap Etch (30mT, 90W, 30sccm CHF3, 10min)
RX	Etch	NEB-31	Aura	RX Resist Etch (Recipe #8)
RX	Etch	Si	PT-720	RX Silicon Fin Etch (30mT, 250V, 97sccm Cl2, 2sccm BCl3, 2.5min)
RX	Rinse	DI	Wet	Rinse in DI Water

Gate	Proc Type	Material or Mask	Tool	Process
RX	Dry	N2	N2 Gun	Blow Dry
PP Gate				
PP	Clean	RCA	MOS	MOS Clean - NO HF Dip
PP	Growth	SiO2	Thermco	Sacrificial Oxidation (Proc 20, 850C, 25min)
PP	Etch	Strip	MOS	17sec HF dip in MOS Clean Hood
PP	Growth	SiO2	Thermco	Gate Oxidation (Proc 20, 850C, 25min)
PP	Dep	Poly	Thermco	Poly Deposition (Recipe 3, 650C, 100min)
PP	I/I	As/BF2	Varian	Gate Pre-Doping Implant
PP	Clean	Acetone/IPA	Spinner	Solvent Clean
PP	Prime	HMDS	YES	Vapor Prime
PP	Spin	1813	Spinner	Spin Photo-Resist (1813, 4000rpm, 60sec)
PP	Bake	1813	HotPlate	Solvent Removal Bake (115C, 60sec)
PP	Expose	CMPFR43	HTG	Contact Alignment and Exposure (3sec)
PP	Bake	1813	HotPlate	Develop (300-MIF, 60sec)
PP	Develop	MIF-312	Wet	Hard Bake (115C, 60sec)
PP	Rinse	DI	Wet	Rinse
PP	Dry	N2	N2 Gun	Blow Dry
PP	Descum	1813	Branson	Descum
PP	Etch	Poly	PT-720	PP Polysilicon Etch (30mT, 250V, 97sccm Cl2, 2sccm BCl3, 5min)
PP	Rinse	DI	Wet	Rinse in DI Water
PP	Dry	N2	N2 Gun	Blow Dry
PP	Strip	1813	Aura	PP Resist Etch (Recipe #8)
PP	Clean	Acetone/IPA	Spinner	Solvent Clean
PC Gate				
PC	CMP	Poly	Strasbaugh	Poly Planarization (IT1400Pad, P1000 Slurry, 5psi, 2x30sec)
PC	Clean	Post-CMP	Hamatech	Post-CMP Clean
PC	Etch	Poly	PT-720	Poly Etchback (30mT, 250V, 97sccm Cl2, 2sccm BCl3, 10sec)
PC	Rinse	DI	Wet	Rinse in DI Water
PC	Dry	N2	N2 Gun	Blow Dry
PC	Dep	Si3N4	Thermco	Deposit Nitride (Hardmask)
PC	Spin	DUV52	CEE6000	Spin DUV ARC (DUV52, 1050rpm, 60sec)
PC	Bake	DUV52	CEE6000	Pre-Bake ARC (DUV52, 205C, 90sec)
PC	Spin	UV-82	CEE6000	Spin DUV Photo-Resist (UV-82, 3000rpm, 20sec)
PC	Bake	UV-82	CEE6000	Pre-Bake Photo-Resist (UV-82, 130C, 60sec)
PC	Expose	NikonFinData	Nikon	DUV Expose (FRIED1.xx)
PC	Bake	UV-82	CEE6000	Post-Bake (140C, 60sec)

Gate	Proc Type	Material or Mask	Tool	Process
PC	Develop	CD-26	CEE6000	Develop (CD-26, 30sec)
PC	Rinse	DI	Wet	Rinse in DI Water
PC	Dry	N2	N2 Gun	Blow Dry
PC	Etch	UV-82	AMAT	Etch ARC, Descum (30mT, 90W, 30sccm O2, 60sec)
PC	Etch	Si3N4	Oxford	PC Nitride Cap Etch (30mT, 65W, 30sccm CHF3, 4min)
PC	Etch	UV-82	Aura	PC Resist Etch (Recipe #8)
PC	Etch	Poly	PT-720	PC Gate Etch (30mT, 250V, 97sccm Cl2, 2sccm BCl3, 4min)
PC	Rinse	DI	Wet	Rinse in DI Water
PC	Dry	N2	N2 Gun	Blow Dry
PC	Etch	SiO2	Wet	Thermal Hardmask Etch (30:1 BHF, 12min)
PC	Clean	RCA	MOS	MOS Clean - NO HF Dip
PC	Growth	SiO2	Thermco	Sidewall Reoxidation (Proc 20, 850C, 25min)
BP/BN Gate				
BP	I/I	As/BF2	Varian	Source/Drain Implant
BP	RTA	1813	AGH	Extension Activation
BP	Dep	SiO2	GSI	Deposit Passivation Oxide (Recipe Undoped Oxide, n=1.46, 2min)
CA Gate				
CA	Prime	HMDS	YES	Vapor Prime
CA	Spin	1813	Spinner	Spin Photo Resist (1813, 4000rpm, 3sec acc, 60sec spin, 3sec dec.)
CA	Bake	1813	HotPlate	Solvent Removal Bake (115C, 60sec)
CA	Expose	FINTE48	GCA 5X	G-Line Litho (Dose = 0.5sec, Focus = 250)
CA	Develop	300-MIF	Wet	Develop (300-MIF, 60sec)
CA	Rinse	DI	Wet	Rinse
CA	Dry	N2	N2 Gun	Blow Dry
CA	Bake	1813	HotPlate	Hard Bake (115C, 60sec)
CA	Descum	1813	Branson	Descum
CA	Etch	SiO2	AMAT	Etch SiO2 (30mT, 90W, 30sccm CHF3, 18min)
CA	Strip	1813	Aura	Resist Strip (Recipe #8)
CA	Clean	Actone/IPA	Spinner	Solvent Clean
CP Gate				
CP	Prime	HMDS	YES	Vapor Prime
CP	Spin	1813	Spinner	Spin Photo Resist (1813, 4000rpm, 3sec acc, 60sec spin, 3sec dec.)
CP	Bake	1813	HotPlate	Solvent Removal Bake (115C, 60sec)
CP	Expose	FINTE48	GCA 5X	G-Line Litho (Dose = 0.5sec, Focus = 250)
CP	Develop	300-MIF	Wet	Develop (300-MIF, 60sec)
CP	Rinse	DI	Wet	Rinse
CP	Dry	N2	N2 Gun	Blow Dry
CP	Bake	1813	HotPlate	Hard Bake (115C, 60sec)

Gate	Proc Type	Material or Mask	Tool	Process
CP	Descum	1813	Branson	Descum
CP	Etch	SiO2	AMAT	Etch SiO2 (30mT, 90W, 30sccm CHF3, 18min)
CP	Etch	Si3N4	Oxford	Etch Nitride (55mT, 150W, 50sccm CHF3, 5sccm O2, 5min)
CP	Strip	1813	Aura	Resist Strip (Recipe #8)
CP	Clean	Actone/IPA	Spinner	Solvent Clean
CP	Wet	SiO2	BHF	Native Oxide Etch & Expand CA's (30:1 BHF, 2min)
CP	Sputter	Al	CVC	Sputter Al (DC Magnetron Sputter, 25min)
M1 Gate				
M1	Spin	HMDS	YES	Vapor Prime
M1	Spin	1813	Spinner	Spin Photo Resist (1813, 4000rpm, 3sec acc, 60sec spin, 3sec dec.)
M1	Bake	1813	HotPlate	Pre-Bake Photo-Resist (115C, 60sec)
M1	Expose	FINTE49	GCA 5X	G-Line Litho (Dose = 1.5sec, Focus = 250)
M1	Reversal	NH3	YES	Image Reversal
M1	Expose	Flood	HTG	Flood Expose (60sec.)
M1	Develop	MF-321	Wet	Develop (MF-321, 60sec)
M1	Rinse	DI	Wet	Rinse
M1	Dry	N2	N2 Gun	Blow Dry
M1	Bake	1813	HotPlate	Hard Bake (115C, 60sec)
M1	Descum	1813	Branson	Descum
M1	Etch	Al	PT-720	Etch Al (Aluminum Etch, 6min)
M1	Strip	1813	Aura	Resist Strip (Recipe #8)
M1	Clean	Actone/IPA	Spinner	Solvent Clean
M1	Deposit	SiO2	GSI	Deposit Inter-Layer Dielectric (n=1.46, 2min)
OG Gate				
OG	Prime	HMDS	YES	Vapor Prime
OG	Spin	1813	Spinner	Spin Photo Resist (1813, 4000rpm, 3sec acc, 60sec spin, 3sec dec.)
OG	Bake	1813	HotPlate	Solvent Removal Bake (115C, 60sec)
OG	Expose	FINTE52	GCA 5X	G-Line Litho (Dose = 0.5sec, Focus = 250)
OG	Develop	300-MIF	Wet	Develop (300-MIF, 60sec)
OG	Rinse	DI	Wet	Rinse
OG	Dry	N2	N2 Gun	Blow Dry
OG	Bake	1813	HotPlate	Hard Bake (115C, 60sec)
OG	Descum	1813	Branson	Descum
OG	Etch	SiO2	AMAT	Etch SiO2 (30mT, 90W, 30sccm CHF3, 18min)
OG	Strip	1813	Aura	Resist Strip (Recipe #8)
OG	Clean	Actone/IPA	Spinner	Solvent Clean
OG	Anneal	FG	Thermco	Surface State Anneal (15% Hydrogen, 400C, 20min)
OG	Test	-	Agilent	Final Test

References

1. Yanagiya, N., *et al. 65 nm CMOS technology (CMOS5) with high density embedded memories for broadband microprocessor applications.* in *Electron Devices Meeting, 2002. IEDM '02. Digest. International*. 2002.

2. Nowak, E., *VDD, VT, TOX Trends,* . 2002.

3. Lee, S.J., *et al. High-k gate dielectrics for sub-100 nm CMOS technology*. in *Solid-State and Integrated-Circuit Technology, 2001. Proceedings. 6th International Conference on*. 2001.

4. Davari, B., R.H. Dennard, and G.G. Shahidi, *CMOS scaling for high performance and low power-the next ten years.* Proceedings of the IEEE, 1995. **83**(4): p. 595-606.

5. Krishnan, S. and J.G. Fossum, *Grasping SOI floating-body effects.* Circuits and Devices Magazine, IEEE, 1998. **14**(4): p. 32-37.

6. Fossum, J.G., V.P. Trivedi, and K. Wu. *Extremely scaled fully depleted SOI CMOS*. in *SOI Conference, IEEE International 2002*. 2002.

7. Solomon, P.M., *et al., Two gates are better than one [double-gate MOSFET process]*. Circuits and Devices Magazine, IEEE, 2003. **19**(1): p. 48-62.

8. Ieong, M., *et al. High performance double-gate device technology challenges and opportunities*. in *Quality Electronic Design, 2002. Proceedings. International Symposium on*. 2002.

9. Ieong, M., *et al. Experimental evaluation of carrier transport and device design for planar symmetric/asymmetric double-gate/ground-plane*

CMOSFETs. in *Electron Devices Meeting, 2001. IEDM Technical Digest. International*. 2001.

10. Esseni, D., *et al., An experimental study of mobility enhancement in ultrathin SOI transistors operated in double-gate mode.* Electron Devices, IEEE Transactions on, 2003. **50**(3): p. 802-808.

11. Guarini, K.W., *et al. Triple-self-aligned, planar double-gate MOSFETs: devices and circuits.* in *Electron Devices Meeting, 2001. IEDM Technical Digest. International*. 2001.

12. Lee, J.-H., *et al. Super self-aligned double-gate (SSDG) MOSFETs utilizing oxidation rate difference and selective epitaxy*. in *Electron Devices Meeting, 1999. IEDM Technical Digest. International*. 1999.

13. Yang, I.Y., *et al., Back-gated CMOS on SOIAS for dynamic threshold voltage control.* Electron Devices, IEEE Transactions on, 1997. **44**(5): p. 822-831.

14. Wong, H.Y., K. Shin, and M. Chan. *The gate misalignment effects of the sub-threshold characteristics of sub-100 nm DG-MOSFETs*. in *Electron Devices Meeting, 2002. Proceedings. 2002 IEEE Hong Kong*. 2002.

15. Fried, D.M., *et al. A sub 40-nm body thickness n-type FinFET*. in *Device Research Conference, 2001*. 2001.

16. Hisamoto, D., *et al., A fully depleted lean-channel transistor (DELTA)-a novel vertical ultrathin SOI MOSFET.* Electron Device Letters, IEEE, 1990. **11**(1): p. 36-38.

17. Hisamoto, D., *et al., FinFET-a self-aligned double-gate MOSFET scalable to 20 nm.* Electron Devices, IEEE Transactions on, 2000. **47**(12): p. 2320-2325.

18. Huang, X., *et al., Sub-50 nm P-channel FinFET.* Electron Devices, IEEE Transactions on, 2001. **48**(5): p. 880-886.

19. Kedzierski, J., *et al. High-performance symmetric-gate and CMOS-compatible V_t asymmetric-gate FinFET devices*. in *Electron Devices Meeting, 2001. IEDM Technical Digest. International*. 2001.

20. Kedzierski, J., *et al. Metal-gate FinFET and fully-depleted SOI devices using total gate silicidation*. in *Electron Devices Meeting, 2002. IEDM '02. Digest. International*. 2002.

21. Kedzierski, J., *et al., Extension and source/drain design for high-performance FinFET devices.* Electron Devices, IEEE Transactions on, 2003. **50**(4): p. 952-958.

22. Nowak, E.J., *et al. A functional FinFET-DGCMOS SRAM cell.* in *Electron Devices Meeting, 2002. IEDM '02. Digest. International*. 2002.

23. Rainey, B.A., *et al. Demonstration of FinFET CMOS circuits.* in *Device Research Conference, 2002. 60th DRC. Conference Digest*. 2002.

24. Taur, Y., *Analytic solutions of charge and capacitance in symmetric and asymmetric double-gate MOSFETs.* Electron Devices, IEEE Transactions on, 2001. **48**(12): p. 2861-2869.

25. Solomon, D.P.M., *Physically Based, Mixed Mode Compact Model for the Double Gated FET*, . 2001: Yorktown Heights, NY.

26. Guillaumot, B., *et al. Metal gate and high-k integration for advanced CMOS devices*. in *Plasma- and Process-Induced Damage, 2003 8th International Symposium*. 2003.

27. Hu, S.F., *et al. Single-electron transistor using self-aligned sidewall spacer gates on silicon-on-insulator nanowire.* in *Nanotechnology, 2003. IEEE-NANO 2003. 2003 Third IEEE Conference on*. 2003.

28. Stiffler, S.R., *et al., Oxidation-induced defect generation in advanced DRAM structures.* Electron Devices, IEEE Transactions on, 1990. **37**(5): p. 1253-1287.

29. Bordonaro, G., *On the DUV Sensitivity of NEB-31 Electron Beam Resist*, . 2003.

30. Ganguly, U., *On the CMP Uniformity of Various Wafer-Scale Features*, . 2003.

31. Gokirmak, A., *On the Mechanical Properties of Various Deposited Silicon Nitride Films*, . 2003.

32. Fried, D.M., *et al. A fin-type independent-double-gate NFET*. in *Device Research Conference, 2003*. 2003.

33. Fried, D.M., J.S. Duster, and K.T. Kornegay, *Improved independent gate N-type FinFET fabrication and characterization.* Electron Device Letters, IEEE, 2003. **24**(9): p. 592-594.

34. Fried, D.M., J.S. Duster, and K.T. Kornegay, *High-Performance P-Type Independent Gate FinFETs*, . 2004: Accepted for publication in IEEE Electron Device Letters, April 2004.

35. Liu, Y., *et al.*, *Ideal rectangular cross-section Si-Fin channel double-gate MOSFETs fabricated using orientation-dependent wet etching.* Electron Device Letters, IEEE, 2003. **24**(7): p. 484-486.

36. Emery, T., *Leica VB6 Multiple Pattern Jobfile*, . 2001.