Device Design and Process Technology for Sub-100 nm SOI MOSFET's

by

Andy Wei

B. S., University of Arizona (1994)

S.M., Massachusetts Institute of Technology (1996)

Submitted to the

Department of Electrical Engineering and Computer Science

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September, 2000

[February 2001]

Signature of Author _____

Department of ~~Electrical Engineering~~ and Computer Science, September 13, 2000

Certified by _____

Dimitri A. Antoniadis, Professor, Thesis Supervisor

Accepted by _____

Arthur C. Smith, Chair, Department Committee on Graduate Students

Device Design and Process Technology for Sub-100 nm SOI MOSFET's

by

Andy Wei

Submitted to the
Deparment of Electrical Engineering and Computer Science
on September 13, 2000, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in Electrical Engineering and Computer Science

Abstract

Silicon-on-insulator (SOI) MOSFET's are an attractive alternative to bulk-silicon MOSFET's in the sub-100 nm gate length regime due to improved performance and/or scalability. The SOI layer thickness in SOI MOSFET's can be sized so that the channel is either partially- or fully-depleted of majority carriers. Partially-depleted (PD) SOI MOSFET's are easier to manufacture than fully-depleted (FD) SOI MOSFET's because a thicker SOI film can be used. However, PD-SOI MOSFET's are difficult to design due to floating-body effects. FD-SOI MOSFET's are easier to design and are more scalable than PD-SOI MOSFET's. However, in the sub-100 nm gate length regime, fully-depleted SOI MOSFET's are difficult to manufacture because a bottom gate is required for electrostatic integrity. In this thesis, floating-body effects in PD-SOI MOSFET's and process technology for fabrication of double-gate FD-SOI MOSFET's will be investigated.

Floating-body effects in PD-SOI MOSFET's result from floating-body voltage modulation by AC and DC changes in source, drain, gate, and substrate terminal voltages. This modulation of the floating-body voltage results in modulation of the drain current. AC modulation of the floating-body voltage occurs through capacitive coupling to rapidly switching terminal voltages. DC modulation of the floating-body voltage occurs through diode currents and impact ionization. Because the time constants of these processes are very different, the DC and AC *I-V* behavior of floating-body PD-SOI MOSFET's are very different from each other, as well as very different from body-contacted SOI MOSFET's or bulk MOSFET's with the body tied to a fixed voltage. Floating-body effects on *I-V* behavior, how they are important in CMOS digital operation, and how they are affected by device design are described and modeled.

Another consequence of floating-body behavior is history dependence in *I-V* behavior. Rapid switching of terminal voltages induce nearly-negligible changes in body majority carrier content since body majority carriers are trapped by body-source/drain junction diodes. However, the change in body majority carrier content can become significant over many switching cycles, eventually reaching a "switching-steady-state" value if kept switching (different value for different switching patterns), and can return to the initial value if the terminal voltages are returned to the initial settings and enough time is allowed to reach equilibrium. This "hysteretic" variation of the

body majority hole content is problematic because device *I-V* behavior changes with changing body majority carrier content. Device design to minimize hysteretic behavior will be presented.

Once the hysteretic *I-V* behavior of floating-body PD-SOI MOSFET's has been understood, floating-body PD-SOI CMOS technology can be optimized to maximize drive current while minimizing hysteresis. This optimized PD-SOI technology can then be compared to bulk CMOS technology. This was done using the 2-D numerical simulator MEDICI in a optimization framework based on the International Technology Roadmap for Semiconductors for the 100 nm-, 70 nm-, and 50 nm-technology nodes.

And finally, double-gate fully-depleted SOI MOSFET's technology was explored. A method to fabricate bury a bottom-gate under a FD-SOI MOSFET was developed based on a flip-bond-transfer technique. This technique is based on chemical-mechanical-polish and wafer fusion bonding. Double-gate SOI MOSFET's were fabricated based with this technique, and a self-alignment scheme for alignment of the bottom-gate which to the top-gate was explored.

Thesis Supervisor: Dimitri A. Antoniadis
Title: Professor

Acknowledgments

I would like to thank Professor Dimitri Antoniadis for his guidance, technical and otherwise, throughout my graduate school career. The intellectual freedom he affords and encourages has made my time here very productive, both personally and professionally.

I would like thank all the fellow students in Dimitri's research group for all the help, knowledge, and assorted truisms passed back and forth over the years -- in order of appearance: Lisa Su, Jarvis Jacobs, Melanie Sherony, Hang Hu, Keith Jackson, Isabel Yang, Mark Armstrong, Tony Lochtefeld, Hasan Nayfeh, Zachary Lee, Ilia Sokolinski, Andy Ritenour, Ihsan Djomehri, and Isaac Lauer. I'd like to thank Ihsan Djomehri for generating the 2-D doping profiles used in the simulations, Tom Langdo for silicon selective epitaxial growth, and Cristina Gordy and Eda Munro for help with administrative matters.

I would like to thank the entire MTL staff, past and present, for all their help in the fab. In particular, Joe Walsh for all his help with etching and CMP, Bernard Alamariu for help with diffusion, deposition, and other assorted philosophical quandaries, and Paul Tierney for his help with mask-making and photolithography. The staff at MIT Lincoln Laboratories were also great. I'd like to thank Andy Loomis for help developing the wafer bonding process, and Skip Hoyt for wafer grinding.

And, above all, it's the friends I've made here that have made the time here that much more enjoyable. Here's to the Miracle of Science (MOS)/Cambridgeport Lecture Series/Friday Night Study Group over the years: Mark, Tony, Ritenour, Fanman, Jim, Isaac, Fletch, Mitch, Kymissis, Arthur Guinness, and Austin Nichols.

Finally, I thank my parents for their support and encouragement through the years. This one's for you.

Project support for this thesis was provided by the Semiconductor Research Corporation, and by the Semiconductor Research Corporation Graduate Fellowship Program.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

The main advantage of silicon-on-insulator (SOI) CMOS technology versus conventional bulk-silicon CMOS technology is the potential for higher speed operation at the same power or lower power dissipation at the same speed [1]. These advantages are more important as scaling of CMOS technologies becomes more difficult to implement. However, on the transistor- and circuit-level, SOI MOSFET's behave differently than bulk-silicon MOSFET's. These differences in device operation and subsequent differences in device design and scaling are studied in this thesis.

1.1 Silicon-on-Insulator (SOI) MOSFET's

SOI MOSFET's are fabricated on a film of single-crystalline silicon which is separated from the silicon substrate by a layer of silicon dioxide. Figure 1-1 compares a cross sectional representation of bulk MOSFET's to SOI MOSFET's. The buried oxide layer, in conjunction with device isolation, fully isolates the SOI MOSFET with dielectric material.

Three common techniques are used to form the buried oxide layer in SOI wafers. The first is the bond and etchback technique (BESOI) illustrated in Figure 1-2 [2]. A bulk-silicon wafer with

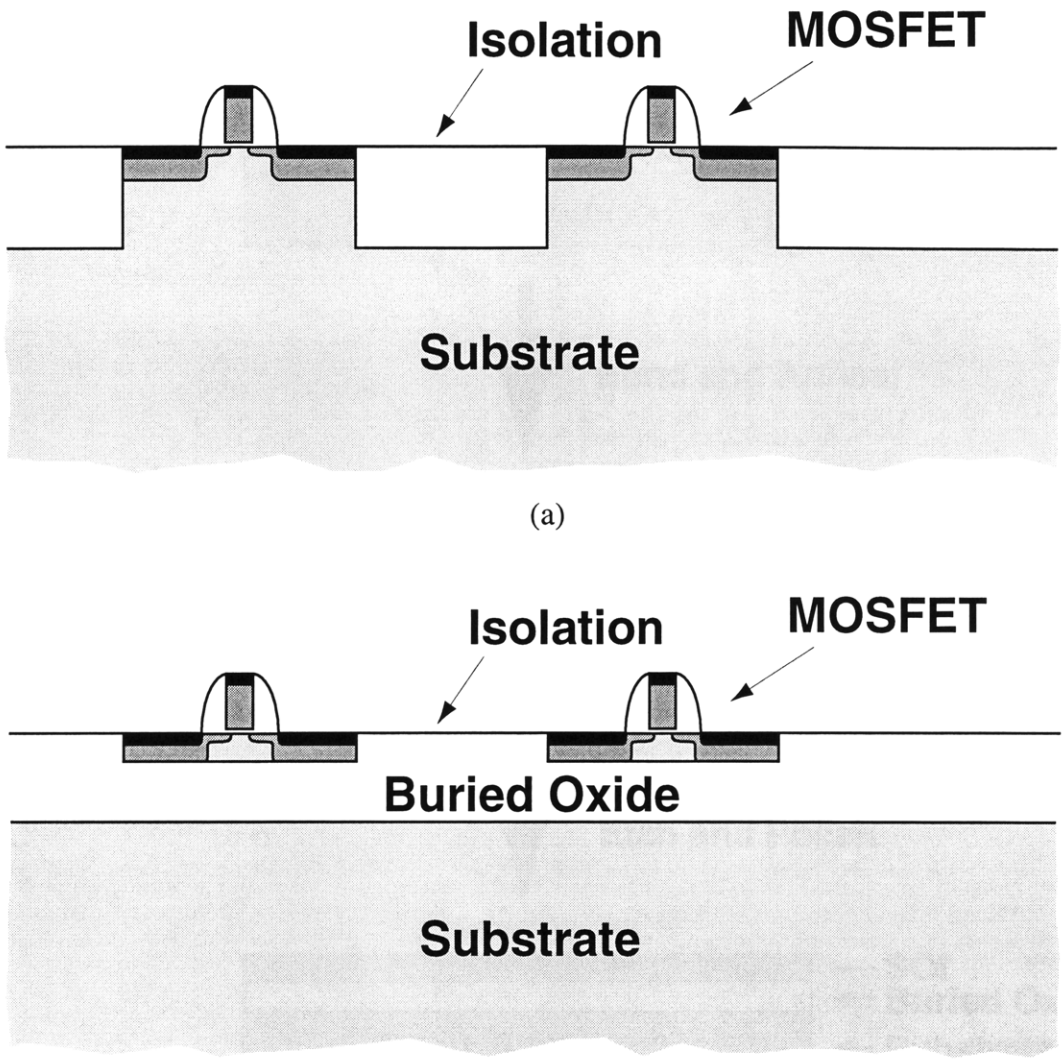


Figure 1-1. Schematic cross section of MOSFET's fabricated on (a) bulk-silicon and (b) silicon-on-insulator, SOI.

thermally-grown oxide is fusion bonded to a bare bulk-silicon wafer. One wafer is etched back and polished to the desired silicon film thickness. This process can produce a large range of buried oxide and silicon film thicknesses, but at the cost of two silicon wafers. The second common techniques involves bond and delaminated, in which a piece of one bulk-silicon wafer is transferred to another wafer with an oxide on top. Such processes include Canon's ELTRAN process [3] and SOITEC's Smart-Cut process [4]. The Smart-Cut process is illustrated in Figure 1-3. Again, two bulk-silicon wafer are required, one bare and one with thermally-grown oxide. One wafer has

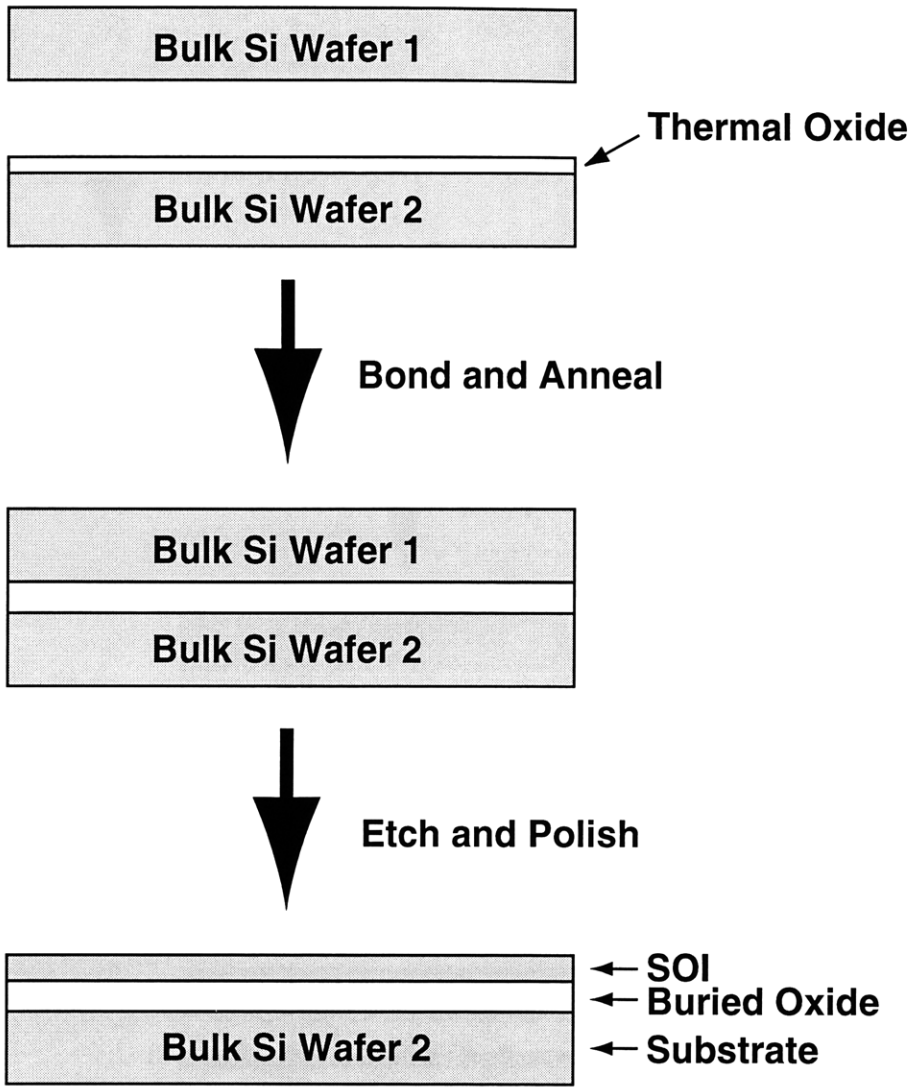


Figure 1-2. Bond and Etchback SOI (BESOI) process for fabricating SOI wafers. Two bulk-silicon wafers, one of which has a thermally-grown oxide, are fusion bonded. The SOI layer is formed by a combination of mechanical grinding, chemical etchback, and mechanical polish.

hydrogen ions implanted at a depth approximately equal to the desired silicon film thickness. The two wafers are bonded at room temperature. The wafer pair is then annealed at a temperature where the expansion of hydrogen into gas fractures the bonded pair at the hydrogen implant peak. It is then annealed at high temperature to complete the fusion bond. The surfaces at the point of fracture are rough and require polishing. After polishing, the bonded pair of wafers produces one SOI wafer and one bulk wafer. In principle, the bulk wafer can be recycled to process more Smart-Cut wafers. Like BESOI, this process can produce a large range of buried oxide thicknesses but

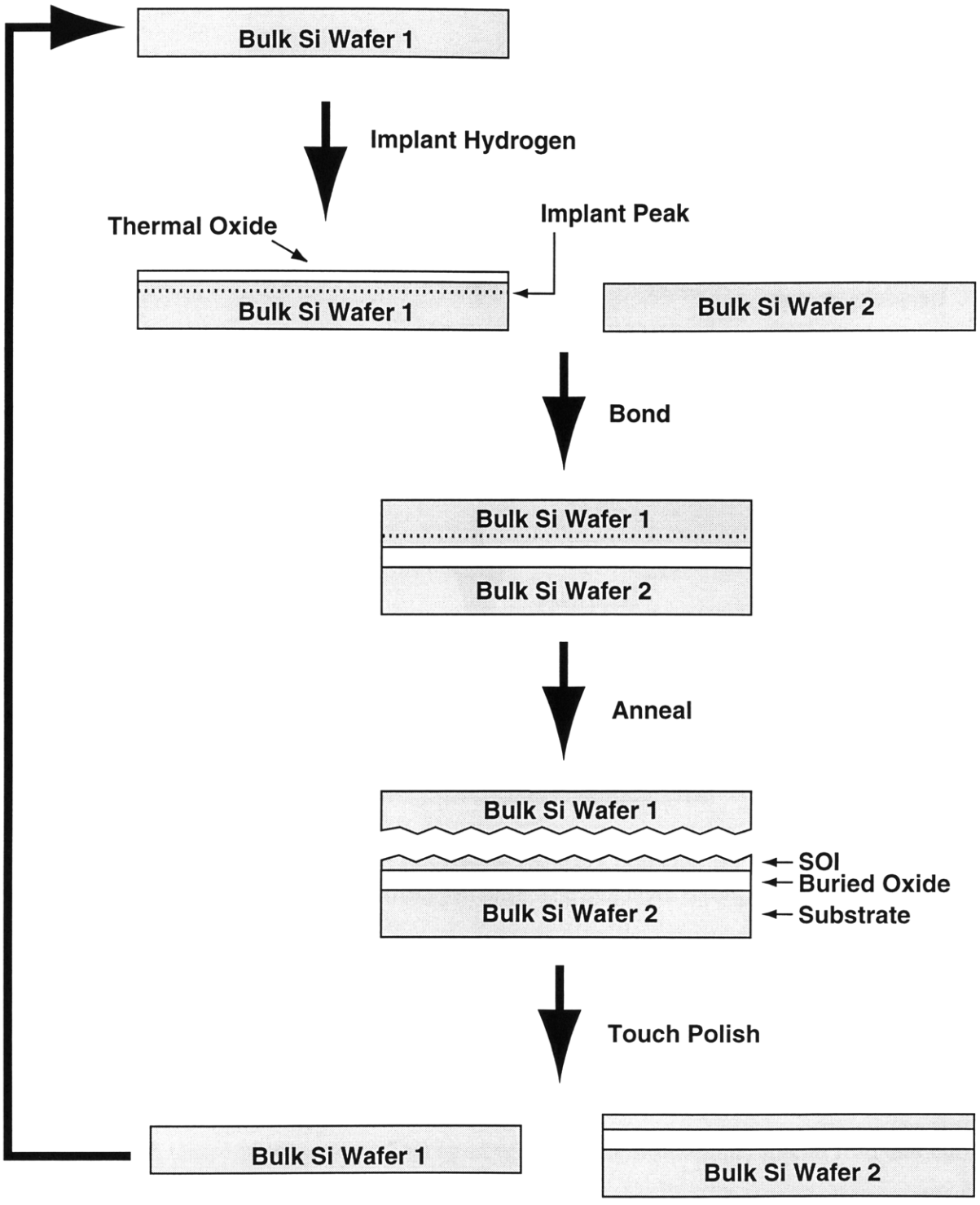


Figure 1-3. Smart-Cut process for fabricating UNIBOND SOI wafers. One bulk-silicon wafer is implanted with hydrogen ions. The SOI layer is formed by fusion bonding the wafer to another wafer with a thermally-grown oxide, at which time the implanted wafer fractures at the hydrogen implant peak. After touch polish of the rough fractured surface, this process produces a SOI wafer and a bulk wafer. The bulk wafer can be recycled to fabricate more SOI wafers.

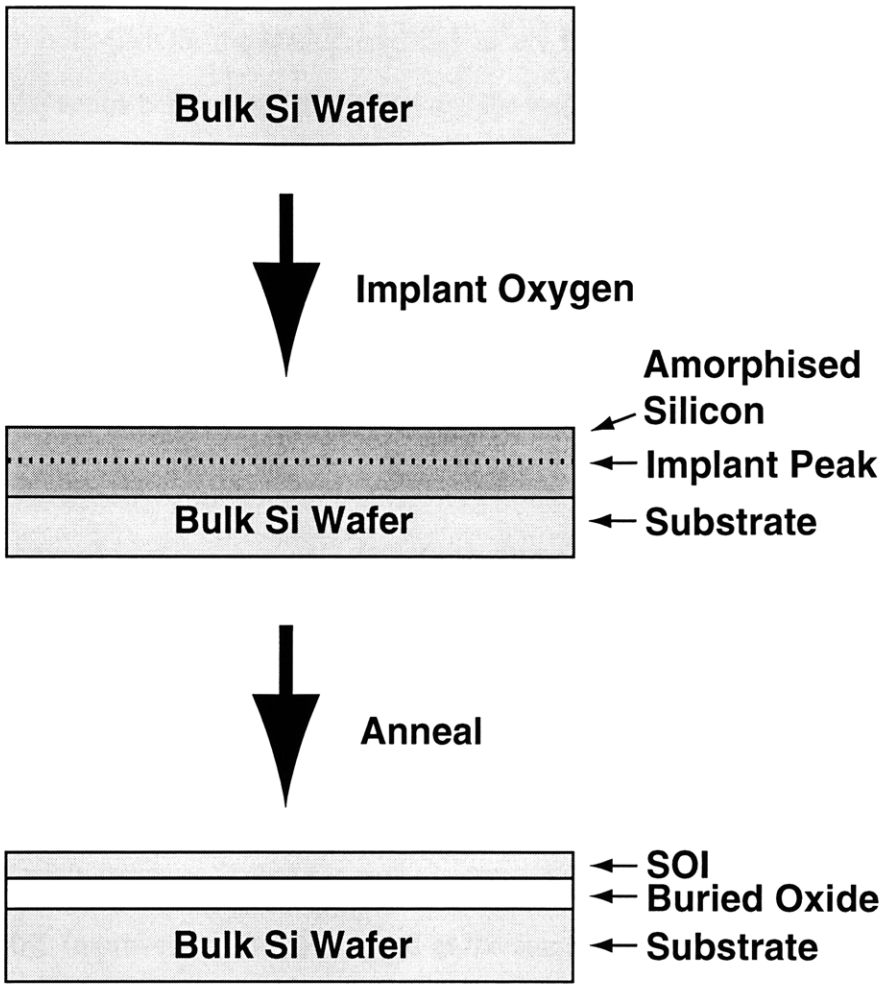


Figure 1-4. Separation by Implantation of Oxygen (SIMOX) process for fabricating SOI wafers. One bulk-silicon wafers is implanted with oxygen ions. The SOI layer is formed by thermal treatment.

silicon film thicknesses are limited by the maximum depth of hydrogen implantation. The third common process, Separation by Implantation of Oxygen (SIMOX), uses only one silicon wafer [5]. As shown in Figure 1-4, the buried oxide layer is formed by a high energy high dose oxygen implant, and high temperature anneal. Due to oxygen implant limitations, silicon film and buried oxide thicknesses are limited [4].

1.2 SOI MOSFET Advantages

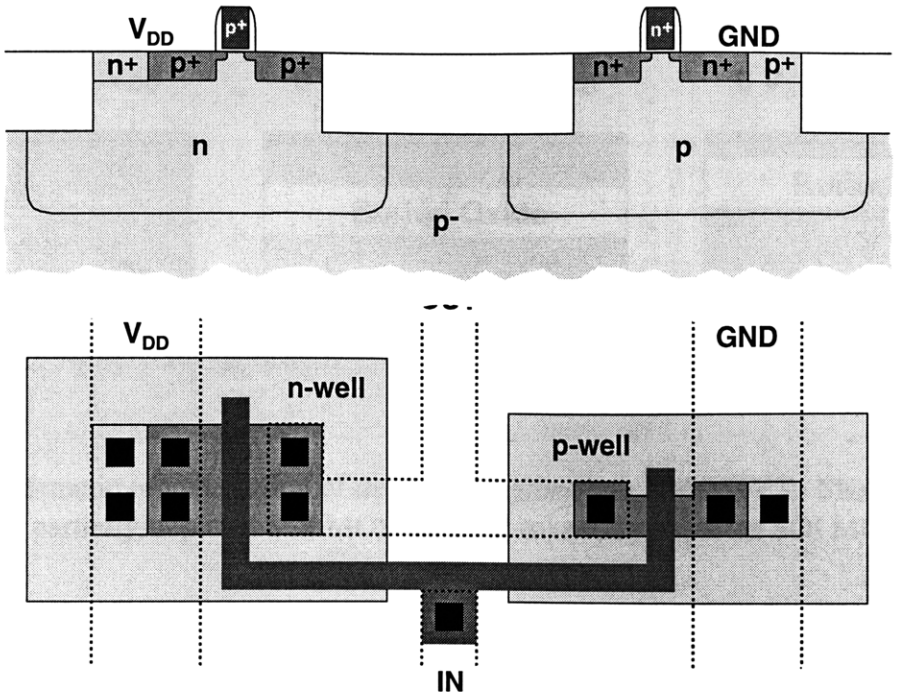
Figure 1-5 compares cross sections and layouts of a CMOS inverter on bulk-silicon to a

CMOS inverter on SOI. In bulk CMOS, the NMOS and PMOS are built in a p-well and n-well, respectively. Parasitic bipolar action between the transistors and the wells can cause a race condition known as latch-up. The transistors and wells must be appropriately spaced and doped to reduce the probability of latch-up. SOI MOSFET's are fabricated on islands of silicon in oxide and thus do not require separate wells for NMOS and PMOS. The NMOS and PMOS can be built in the same silicon island because the drains of the NMOS and PMOS are at the same potential. Thus SOI potentially offers reduced layout complexity, process complexity, and increased packing density [6].

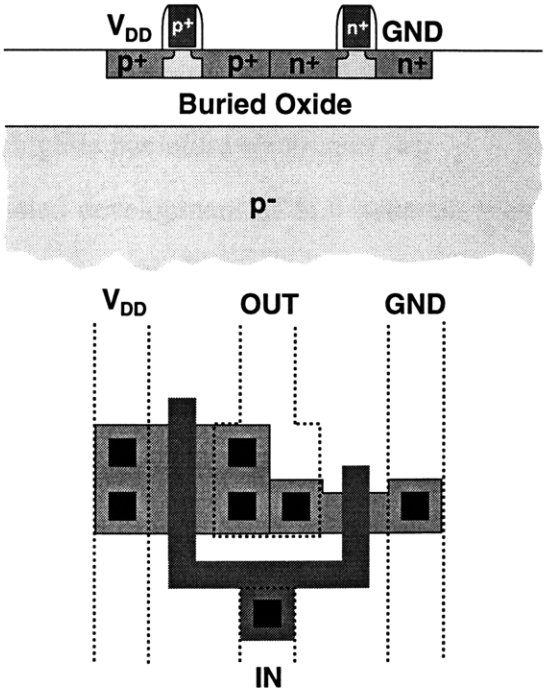
SOI also offers increased performance on a transistor- and circuit-level. The increased performance is due to much reduced parasitic capacitance between the source/drain and the substrate. Further performance gains can be obtained when the body of the SOI MOSFET is floating. This results in increased drive current due to capacitive coupling to the gate [7]. The floating body also improves pass-gates and logic gates with stacked transistors by reducing the instances of reverse body-source biases [8]. It has been demonstrated that the increased performance of SOI CMOS translates to as much as 30% improvement in CPU speed at the same power consumption of an equivalent bulk CMOS CPU, or a 50% decrease in power consumption at the same speed of an equivalent bulk CMOS CPU [1].

1.3 Historical Perspective

SOI has not been adopted for mainstream CMOS digital logic until just prior to the writing of this thesis [9]. There was a reluctance to adopt SOI technology for mainstream CMOS applications because the performance benefits mentioned above could be obtained by scaling bulk CMOS one generation. Given the historic ease of scaling bulk CMOS, redesign of circuit libraries to take full advantage of SOI as well as the increased substrate cost of SOI could not be justified. As scaling of CMOS has become more difficult due to material and lithography limitations, the potential advantages of SOI are now worthy of consideration for mainstream products.



(a)



(b)

Figure 1-5. (a) Cross section and corresponding layout of a CMOS inverter built on bulk-silicon in a twin-well process. (b) Cross section and corresponding layout of a CMOS inverter built on SOI. Dimensions are drawn to scale.

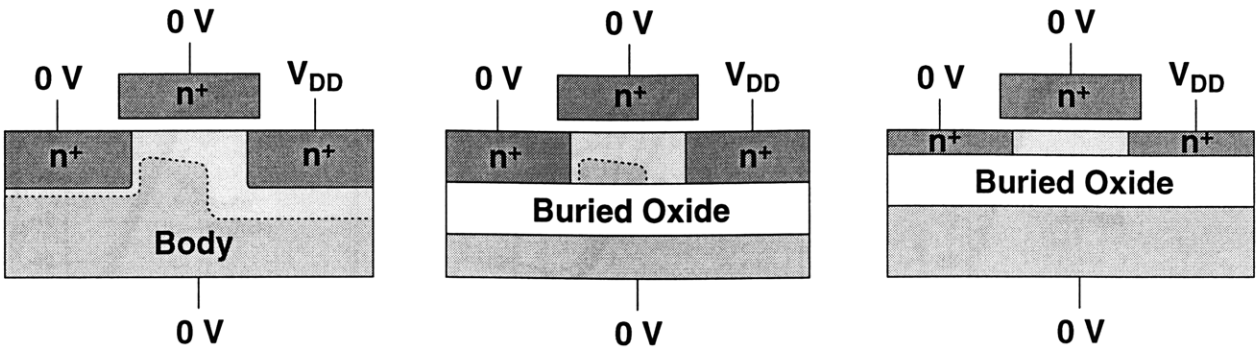


Figure 1-6. Schematic representation of depletion regions in a NMOSFET. Shown is (a) a bulk MOSFET, (b) a partially-depleted SOI MOSFET, and (c) a fully-depleted SOI MOSFET.

Prior to current adoption of SOI for mainstream CMOS production, SOI had been manufactured only for niche applications such as radiation-hard electronics and high-temperature electronics. The full dielectric isolation of SOI MOSFET's reduces single-event upsets by shielding the channel region from excess electron-hole pairs generated by alpha particle penetration into the substrate [6]. Similarly, full dielectric isolation limits the source/drain junction area thereby limiting junction leakage currents in high-temperature electronics [6].

Such applications fueled development of SOI material, which in the 1960's and 1970's included techniques such as heteroepitaxial growth of silicon film on a single-crystalline insulator (silicon-on-sapphire, SOS), laser or e-beam recrystallization of polysilicon on oxides (zone-melt-recrystallization, ZMR), or homoepitaxial growth of silicon on oxides (epitaxial lateral overgrowth, ELO). These techniques usually resulted in less than perfect silicon films. SOI MOSFET's thus had inferior performance compared to bulk MOSFET's due to inferior mobilities, high interface trap density at the silicon-insulator interface, high dislocation densities, stacking faults, or grain boundaries. Because of these limitations, SOI MOSFET's were only manufactured for niche applications where bulk MOSFET's could not be used.

In 1977, Izumi, et al., proposed the SIMOX technique of creating the buried oxide layer [5]. Subsequently, the BESOI technique was developed and in recent years the Smart-Cut and ELTRAN

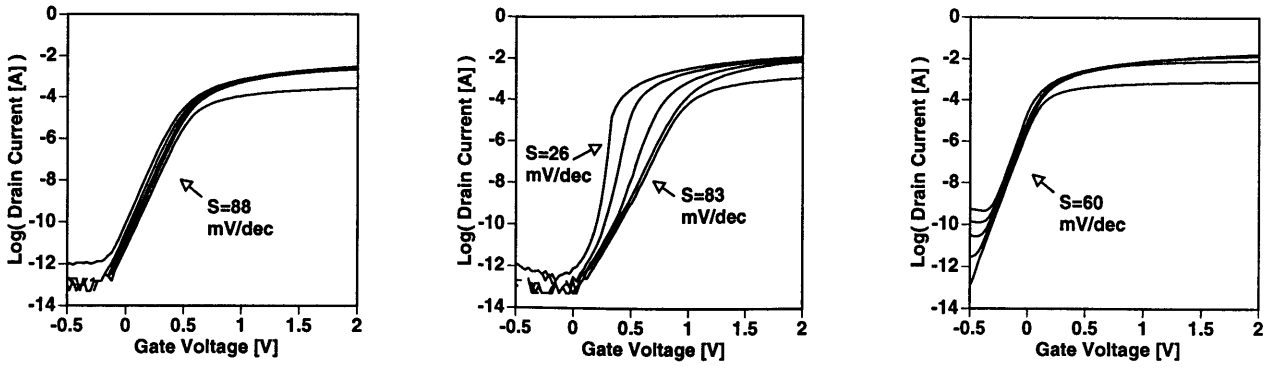


Figure 1-7. Examples of NMOS DC drain current versus gate voltage characteristics for (a) bulk MOSFET's, (b) partially-depleted SOI MOSFET's, and (c) fully-depleted SOI MOSFET's. The *I-V*'s shown represent values of I_{DS} for V_{DS}=0.05 V to 2.05 V by 0.5 V steps. The subthreshold swing, *S*, is indicated on each plot.

techniques were developed. All these techniques produce SOI wafers with a high quality silicon film which can yield ULSI electronics on par with bulk-silicon wafers [10]. Thus, SOI CMOS for mainstream ULSI applications became a reality, and has become competitive with bulk-silicon CMOS as scaling CMOS technologies becomes more difficult. However, SOI CMOS is not without its problems. Differences in how the SOI MOSFET's operate and scale compared to bulk MOSFET's result in different transistor- and circuit-level design issues.

1.4 SOI MOSFET Design Issues

The buried oxide layer truncates the depth of the MOSFET built on SOI. Thus, SOI MOSFET's can have channels which are either fully depleted or partially depleted of majority carriers, depending on body doping and silicon film thickness. This concept is schematically shown in Figure 1-6 for NMOSFET's. Figure 1-7 shows sample DC subthreshold gate *I-V* characteristics for a bulk, partially-depleted (PD) SOI, and fully-depleted (FD) SOI NMOSFET.

1.4.1 Partially-Depleted SOI MOSFET's: Floating-Body Effects

The buried oxide layer leaves the body of the SOI MOSFET floating. While this can be a performance benefit, as noted above, the floating body also introduces anomalous device behavior. For example, the "kink" in the DC subthreshold *I-V* characteristics of the PD-SOI NMOSFET in Figure 1-7 (b) is due to impact-ionization-generated holes which are trapped in the floating quasi-neutral body region by the potential barrier of the body-source p-n junction [11]. Impact-ionization-generated excess holes raise the body voltage which lowers the threshold voltage.

All other "floating-body effects" arise from the same principle of majority carriers trapped in the body by the potential barrier between the body and source/drain [12-19]. Rather than freely moving in and out of the body through the body contact, the only path for majority carriers in and out of the floating body is through forward bias p-n junction conduction, reverse bias p-n junction space charge region generation, and impact ionization generation. The time constants of these paths are much slower compared to device switching timescales. Thus, when the terminal voltages are changed instantaneously (AC), the body majority carrier content cannot change instantaneously. Over time, currents flowing through the body-source/drain junction diodes and impact ionization generation will change the body majority carrier content as the device settles towards equilibrium (DC) behavior [20].

This difference between AC and DC behavior results in hysteretic behavior, where the instantaneous MOSFET behavior depends on the switching history of the device. For example, the threshold voltage can start from equilibrium at one value and change over several switching cycles. This makes the floating-body PD-SOI MOSFET difficult to characterize, model, and design. The body of the PD-SOI MOSFET can be contacted, as shown in Figure 1-8. Body-contacted PD-SOI MOSFET's behave exactly like a bulk MOSFET. However, for the same device width, a body contact consumes extra layout area and increases process complexity. So, while body contacts may be selectively used in parts of circuits which cannot tolerate hysteretic behavior, a majority of

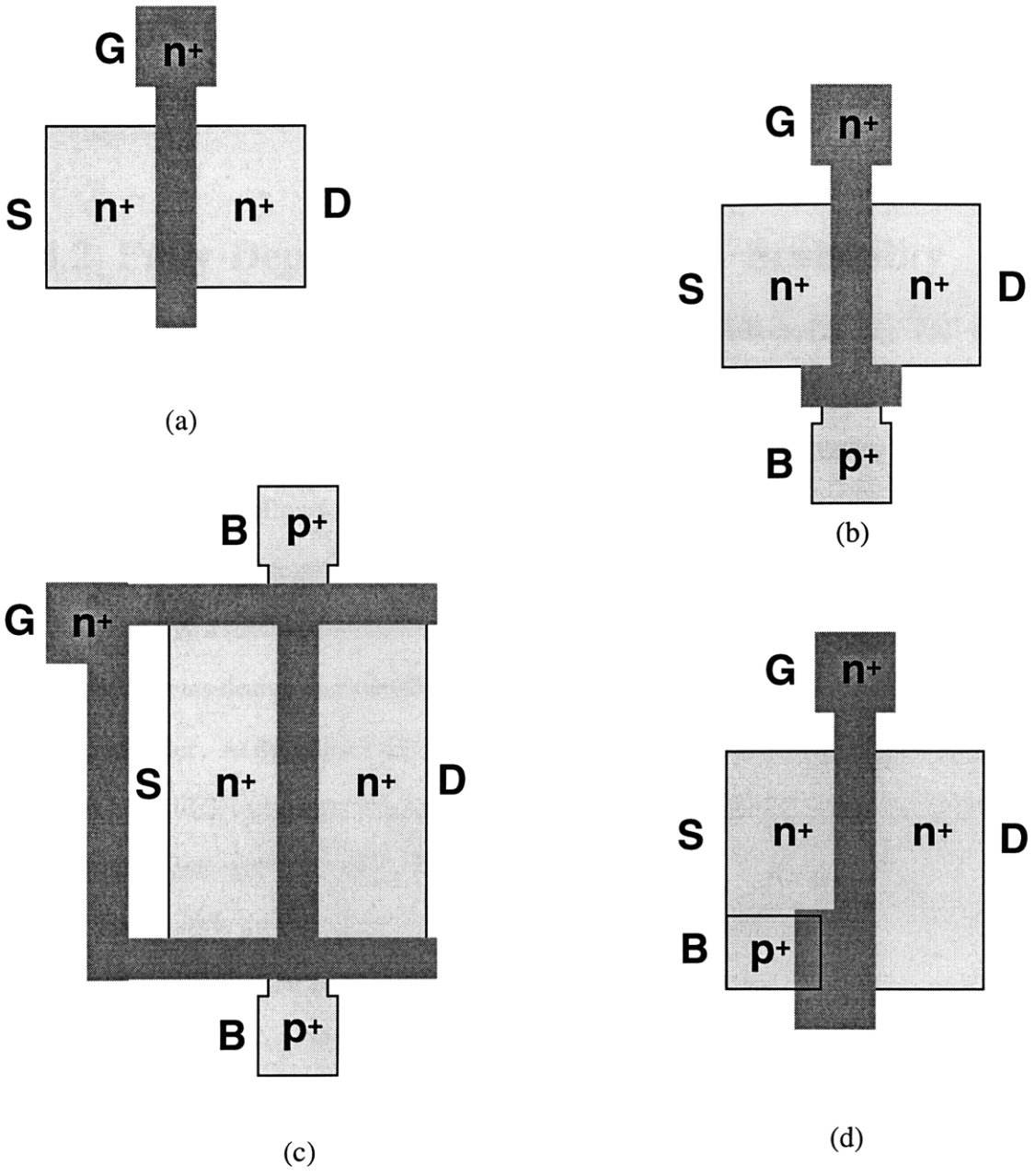


Figure 1-8. Comparison of floating-body and body-contacted SOI MOSFET layouts. (a) Floating-body. (b) Body-contacted. Due to high body resistance, the body contact may become too resistive with increasing device width. Two body contacts are required for wider devices as shown in (c). (c) Asymmetric body tie with body tied to source. The source and drain are not interchangeable. For very wide devices, this structure can be extended with multiple body-ties.

devices should operate with the body floating. Thus, floating-body effects in PD-SOI MOSFET's are an important design issue.

1.4.2 Fully-Depleted SOI MOSFET's: Scalability

FD-SOI MOSFET's display minimized floating-body effects [21,22]. Full-depletion of the body significantly reduces the potential barrier between the body and the source, so majority carriers are not as easily trapped in the body. FD-SOI MOSFET's also have a major advantage because the fully-depleted body allows the surface potential to be directly coupled to the gate voltage [23]. This results in a near-ideal 60 mV/dec subthreshold swing as shown in Figure 1-7 (c). Typical bulk MOSFET's have a subthreshold swing of 80 mV/dec. The subthreshold swing of PD-SOI MOSFET's varies with device design and switching history due to floating-body effects, but is typically between 70 to 85 mV/dec. At the same OFF-state leakage current, improved subthreshold swing allows the FD-SOI MOSFET to operate at a lower threshold voltage, V_T, and achieve increased drain current due to higher gate overdrive $(V_{DD}-V_T)$.

Thus, while a FD-SOI MOSFET is superior to bulk and PD-SOI MOSFET's, it is difficult to scale and manufacture because thin silicon films are required to maintain full depletion of the body. Obviously, a FD-SOI MOSFET without a sufficiently thin silicon film is a PD-SOI MOSFET with all the associated floating-body effects. Maintaining full-depletion of the body is achieved by reducing the silicon film thickness, as shown in Figure 1-9 and can be described as follows: An ideal transistor behaves as a gate-controlled current source. A well-behaved MOSFET should have saturation drain current which is greatly affected by changes in the gate voltage (high transconductance, $g_m=\partial I_{DS}/\partial V_{GS}$), but minutely affected by changes in the drain voltage (low output conductance, $g_o=\partial I_{DS}/\partial V_{DS}$). The level of gate control versus drain control of the drain current is directly related to the relative level of gate control to drain control of body majority carriers and channel charge. For a long channel device with a thin silicon film, the gate controls most of the channel charge compared to the drain. As the channel becomes shorter, the portion of channel

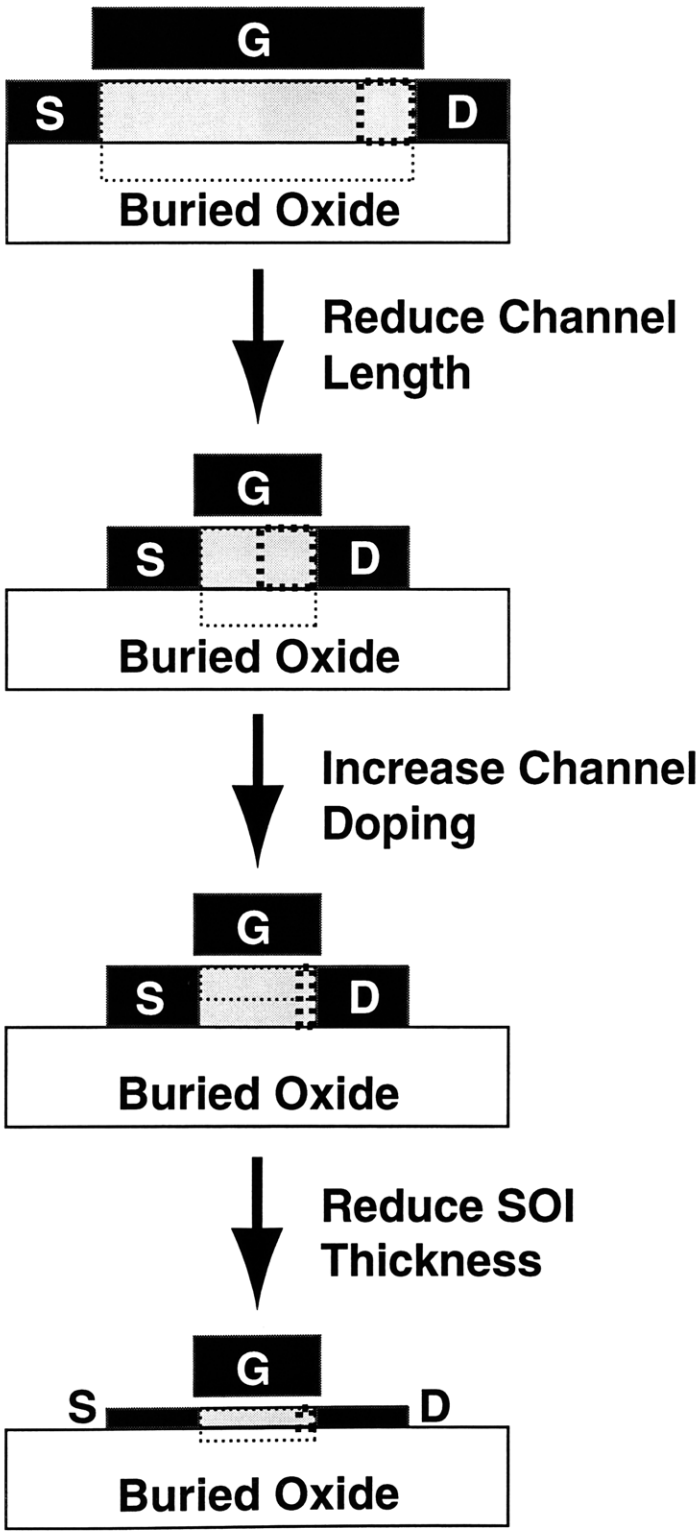


Figure 1-9. Simplified schematic demonstrating fully-depleted SOI film thickness scaling. Shown in the light dashed line is charge controlled by the gate, and in the bold dashed line charge controlled by the drain. Reduction of SOI film thickness is necessary to maintain full depletion of the channel.

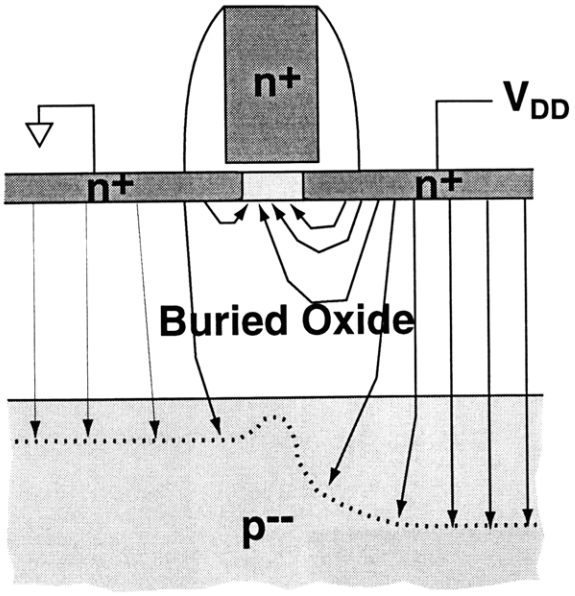


Figure 1-10. Schematic representation of 2-D channel charge sharing through the buried oxide in fully-depleted SOI NMOSFET.

charge controlled by the drain versus the gate increases. To counteract this "short-channel effect", the channel doping is increased. This reduces the depletion depth from the gate. As a result, the silicon film thickness must be reduced when the channel length is reduced in order to maintain a fully-depleted channel.

Now, as the gate length is reduced, maintaining a suitable threshold voltage requires higher channel doping densities which in turn requires thinner silicon films in order to achieve full depletion. This is problematic because high channel dopings reduce carrier mobility and thin silicon films are difficult to isolate and contact. In addition, these problems are exacerbated by drain control of channel charge through the buried oxide region, as shown in Figure 1-10 [24]. The field lines from the drain favorably terminate on channel charge rather than on substrate charge due to lower dielectric constant of oxide, and low substrate doping. This "2-D charge-sharing" effect increases drain control of channel charge versus gate control which forces the use of even higher channel doping and even thinner silicon film.

These scaling problems can be alleviated by reducing the buried oxide thickness and inserting a highly doped substrate region beneath the channel region [25]. This highly doped substrate

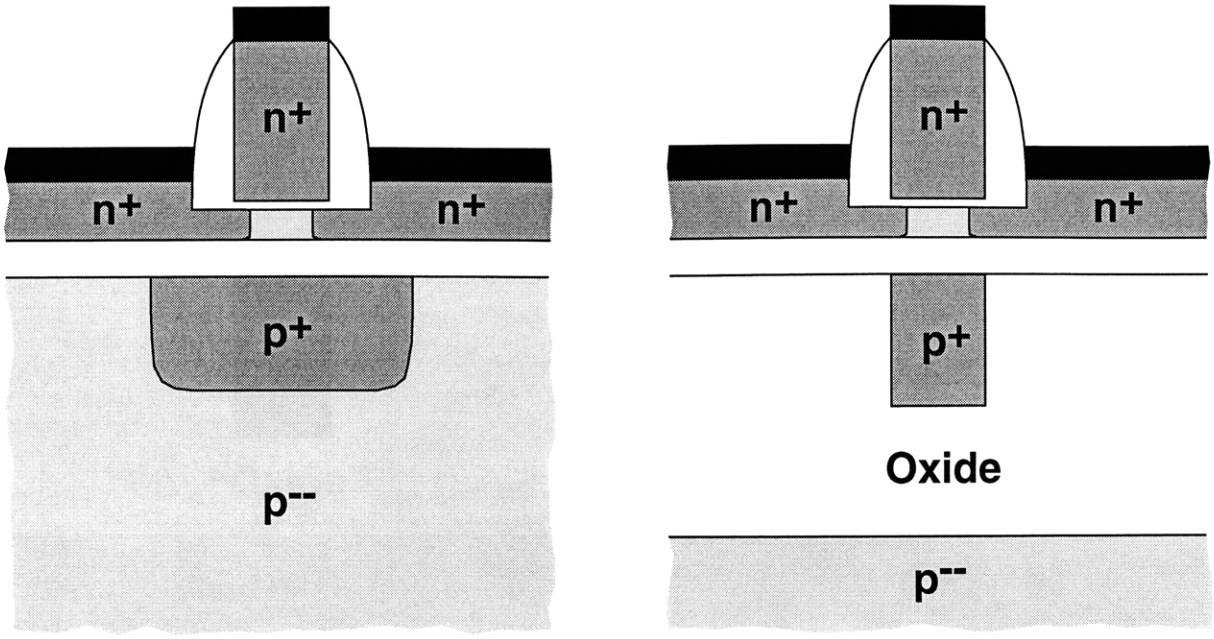
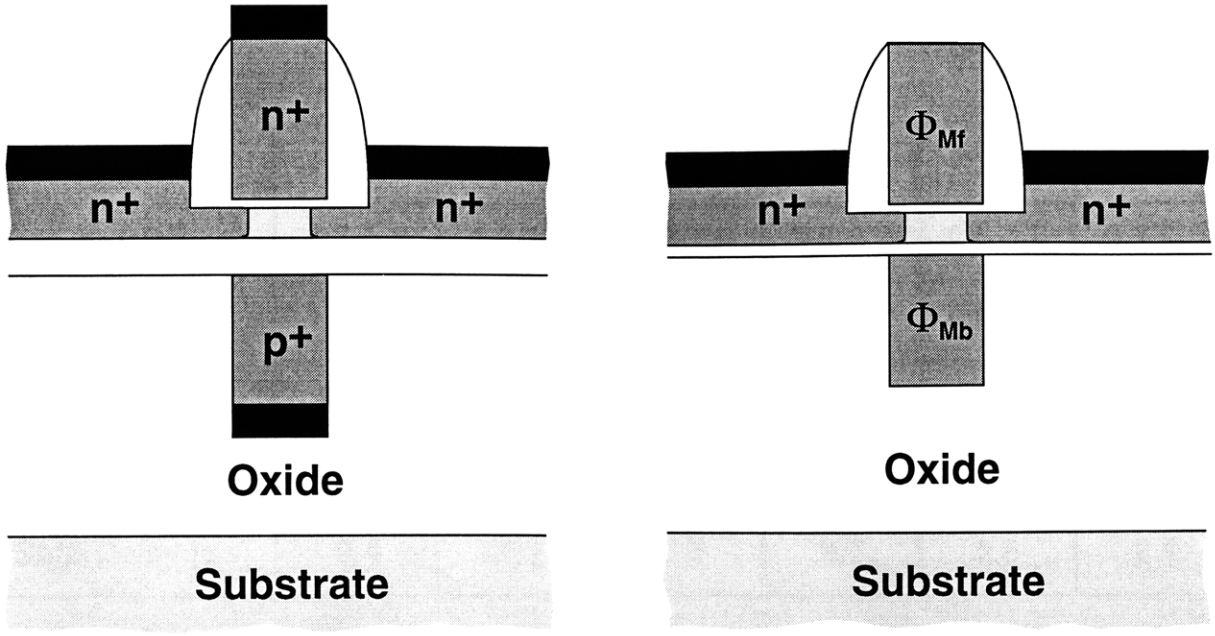


Figure 1-11. Schematic representation of (a) a scaled fully-depleted SOI NMOSFET with thin buried oxide, highly doped substrate region, and salicided raised source/drain. Parasitic capacitances can be further reduced with the structure shown in (b).

region also reduces the channel doping level because of the workfunction difference. A raised source drain can be grown epitaxially to form good electrical contact to the source and drain, and thus low parasitic series resistance [26]. A schematic of an optimized short-channel FD-SOI MOSFET structure is shown in Figure 1-11 (a). This structure can be further optimized to reduce the parasitic source/drain-substrate capacitance by replacing the substrate with oxide and aligning the doped region underneath the channel to the top gate, as shown in Figure 1-11 (b).

If the highly doped region under the channel is contacted, the optimized fully-depleted SOI MOSFET becomes a double-gate MOSFET. A double-gate MOSFET can exhibit increased performance over a single-gate MOSFET if the bottom-gate is used to modulate the MOSFET threshold voltage so that it is low during active periods and high during idle periods [27]. Such a device can be implemented in an asymmetric gate structure shown in Figure 1-12 (a). Further improvement in device performance can be obtained with a symmetric double gate structure by operating both gates simultaneously on an undoped channel region, where the gate material



(a) (b)

Figure 1-12. Schematic representation of (a) a double-gate fully-depleted SOI NMOSFET with asymmetric gates and (b) with symmetric gates. The gate workfunctions of the symmetric double-gate device are shown as design variables.

workfunction is used to obtain a proper threshold voltage, as shown in Figure 1-12 (b).

It is difficult to fabricate two gates well-aligned to each other with planar fabrication technology, however. New methods must be developed to place a gate below the a FD-SOI MOSFET in a self-aligned manner.

1.5 Scaling into the Sub-100 nm Regime

To put SOI MOSFET design, fabrication, and scaling into perspective, Table 1-1 shows proposed scaling trends as outlined by the International Technology Roadmap for Semiconductors (ITRS) [28]. Performance is expected to increase based on a decreasing gate delay metric. This is achieved by maintaining drive current at approximately 750 μA/μm NMOS and 350 A/μm for PMOS, while intrinsic capacitances are reduced with shrinking of device dimensions. Meanwhile, scaled transistors maintain electrostatic integrity by decreases in gate oxide thickness and source/

Technology	180 nm	130 nm	100 nm	70 nm	50 nm	35 nm
L_{GATE} [nm]	140	85	65	45	32	22
V_{DD} [V]	1.5-1.8	1.2-1.5	0.9-1.2	0.6-0.9	0.5-0.6	0.3-0.6
$t_{ox,eq}$ [nm]	1.9-2.5	1.5-1.9	1.0-1.5	0.8-1.2	0.6-0.8	0.5-0.6
x_j [nm]	45-70	30-50	25-40	20-28	13-20	10-14
I_{OFF} [nA/μm]	5	10	20	40	80	160
Gate Delay [ps]	9.4	7.3	5.7	3.7	2.6	2.4

Table 1-1. High-performance logic technology requirements from the International Technology Roadmap for Semiconductors.

drain junction depth. Maximum electric fields in the gate oxide and reduced power levels in circuits dictate a decrease in the power supply, V_{DD}, with each generation, and V_T must be decreased (OFF-state leakage, I_{OFF}, increases) to maintain sufficient gate overdrive (V_{DD}-V_T).

1.6 Thesis Goals

PD-SOI MOSFET's are potentially more scalable than bulk MOSFET's if the floating-body effects are properly controlled on a transistor level and properly exploited on a circuit level. FD-SOI, in the guise of a double-gate MOSFET, can potentially scale even better than PD-SOI MOSFET's. However, a double-gate MOSFET is difficult to fabricate. This thesis addresses device design of floating-body PD-SOI MOSFET's and fabrication technology for FD-SOI MOSFET's. While many of the floating-body effects have been reported in [19], this thesis will look into these effects in depth, treating them in the context of digital logic. In **Chapter 2**, the floating body of a PD-SOI MOSFET will be modeled as an equivalent network of diodes, current sources, and

capacitors. The effect of device design on floating-body effects will be demonstrated. In **Chapter 3**, hysteretic floating-body effects will be described. Simulation methodologies which can predict bounds of hysteretic behavior will be demonstrated. Methods to minimize hysteretic floating-body effects will be discussed. With the understanding of floating-body effects, PD-SOI MOSFET's will be compared to bulk-silicon MOSFET's in **Chapter 4**. Optimization to maximize device performance and scalability while minimizing floating-body effects will be studied at sub-100 nm technology nodes based on criteria shown in the 1999 International Technology Roadmap for Semiconductors. In **Chapter 5**, a fabrication technology for double-gate FD-SOI MOSFET's will be demonstrated. **Chapter 6** presents a summary of the main conclusions of this thesis and discusses and recommends topics for future work. **Appendix A** compiles sample MEDICI floating-body simulation input decks, and **Appendix B** lists the steps used to fabricate the double-gate MOSFET presented in Chapter 5.

Chapter 2

The Floating Body

To maximize layout density, the body of the SOI MOSFET is typically left uncontacted and is therefore electrically floating. In a partially-depleted (PD) SOI MOSFET, the floating-body voltage can be modulated by DC and AC changes in source, drain, gate, and substrate terminal voltages. This modulation of the floating-body voltage results in modulation of the drain current. This makes the DC and AC *I-V* behavior of floating-body SOI MOSFET's different from body-contacted SOI MOSFET's or bulk MOSFET's, which have the body tied to a fixed voltage. These differences are collectively called "floating-body effects". This chapter describes floating-body effects in *I-V* behavior, how they are important in CMOS digital operation, and how they are affected by device design. (For simplicity, "bulk MOSFET's" will refer to both bulk and body-contacted SOI MOSFET's, and only NMOSFET's will be described – PMOSFET behavior is similar).

2.1 Bulk vs PD-SOI MOSFET Switching Behavior

Figure 2-1 defines the NMOSFET OFF-state and ON-state with respect to digital logic applications. The OFF-state is defined drain voltage (V_D) at the supply voltage (V_{DD}) and the gate

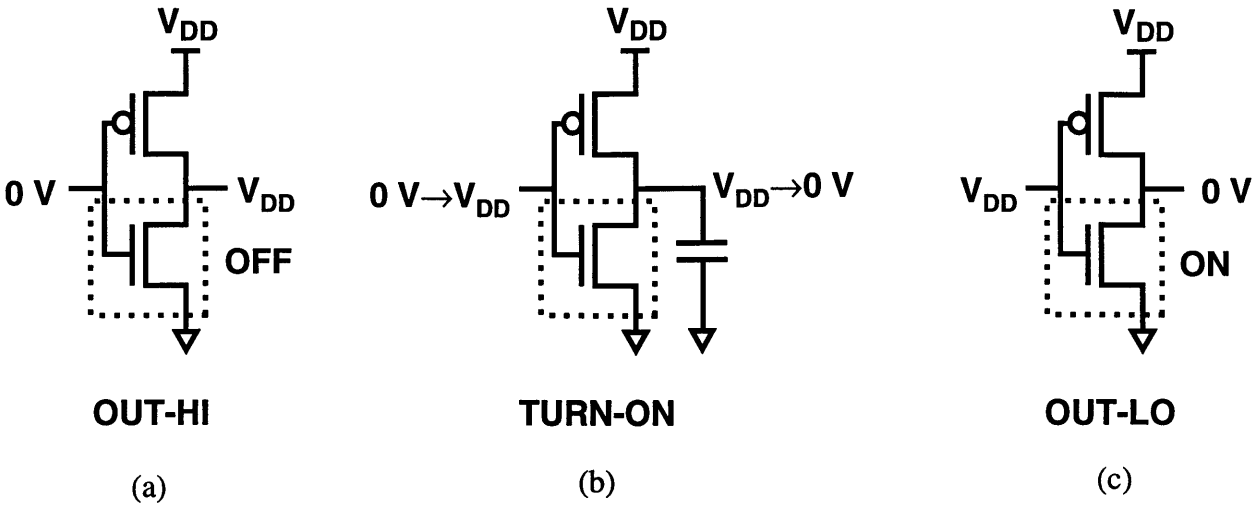


Figure 2-1. Definition of (a) OFF-state and (c) ON-state of a NMOSFET in a CMOS inverter in relation to the logic state OUT-HI and OUT-LO of the CMOS inverter. (b) shows the NMOSFET turning ON with the gate voltage rising and, after a delay due to load capacitance, the drain voltage falling.

voltage at 0 V. The ON-state is defined as V_D=0 V and V_G=V_{DD}. In between these two is the TURN-ON-state where V_G rises from 0 V to V_{DD}. After discharging a load capacitance, V_D falls from V_{DD} to 0 V.

Figure 2-2 shows the currents which flow in a bulk NMOSFET in the OFF-state (V_G=0 V, V_D=V_{DD}, and V_S=V_B=GND). OFF-state subthreshold drain current is made up of electrons in the source diffusing to the drain. These electrons generate electron-hole pairs by impact ionization in the high electric field region of the reverse biased body-drain junction. The generated electrons are instantaneously swept towards the highest potential into the drain contact, and the generated holes are instantaneously swept toward the lowest potential into the body contact. Thermal generation in the reverse-biased body-drain junction depletion region and in the channel region depleted by the gate also generates electron-hole pairs. Again, the electrons are instantaneously swept into the drain and holes are instantaneously swept into body contact.

Figure 2-3 shows the currents which flow when the bulk NMOSFET is turning ON (V_G=0 to V_{DD}, V_D=V_{DD}, and V_S=V_B=GND -- V_D will fall after the NMOSFET discharges a load capacitance).

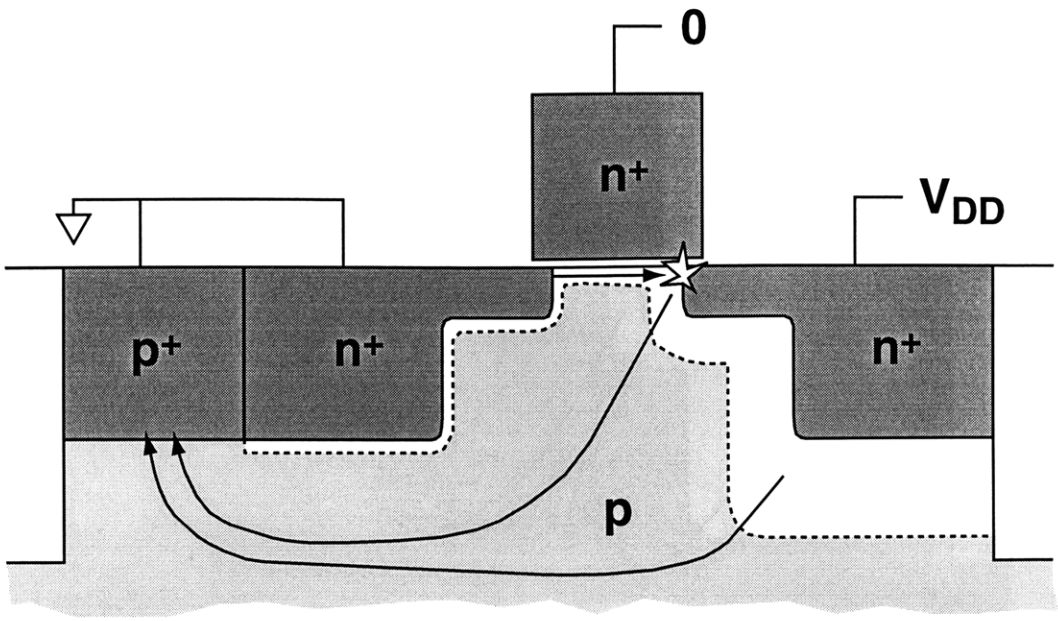


Figure 2-2. Schematic cross section of a bulk NMOSFET in the OFF-state. Arrows show impact ionization current and drain-body depletion region generation current flowing to the body contact.

As the gate voltage rises, the body region depleted by the gate expands until the surface potential favors formation of an inversion layer made of electrons. The excess gate-body depletion-region holes are instantaneously removed from the body through the body contact. TURN-ON-state drain current is made up of inversion layer electrons carried by drift from the source to the drain. These electrons generate electron-hole pairs by impact ionization in the high field region of the reverse biased body-drain junction. Due to much higher drain current, TURN-ON impact ionization generation is much greater than OFF-state impact ionization generation. Thermal generation of electron-hole pairs in the depletion regions is slightly higher than the OFF-state due to increased gate-depletion width. Again, all generated electrons go to the drain and all generated holes are instantaneously removed from the body through the body contact.

Without a body contact, the behavior of the SOI MOSFET is significantly different from the bulk MOSFET. Figure 2-4 shows the currents in a SOI MOSFET in the OFF-state. As in the bulk MOSFET there is subthreshold drain current causing impact ionization generation of electron hole pairs and there is thermal generation of electron-hole pairs in the reverse biased body-drain

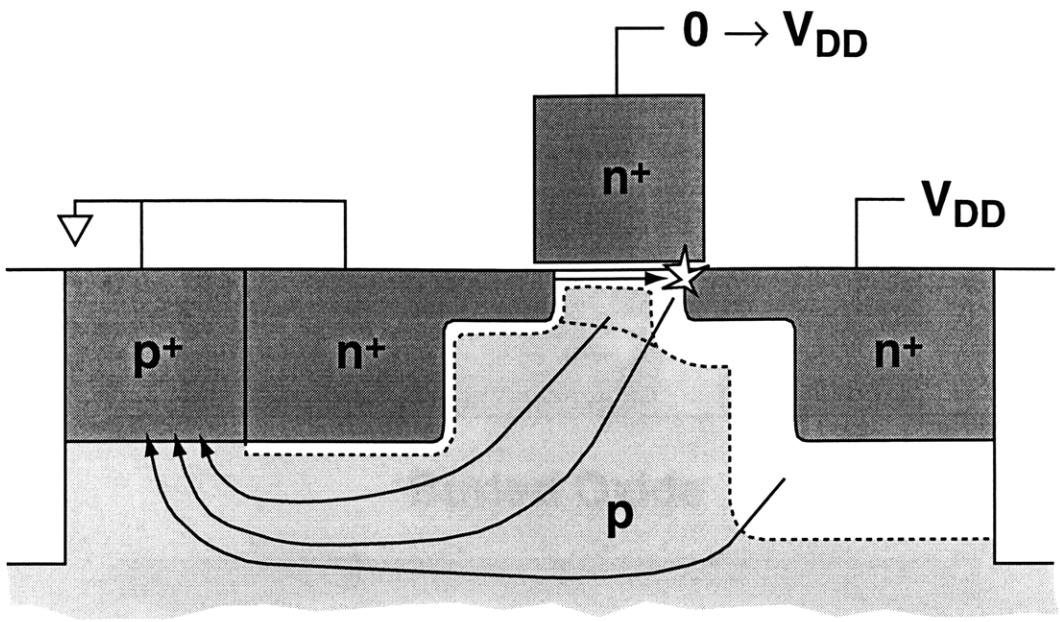


Figure 2-3. Schematic cross section of a bulk NMOSFET switching from the OFF-state to the ON-state. Arrows show gate-depleted body charge, impact ionization current, and drain-body depletion region generation current flowing to the body contact.

junction. The electrons are swept into the drain and the holes are swept into the body. The holes, however, are trapped in the body region because of the potential barrier of the body-source junction. The excess hole content raises the body voltage, forward biasing the body-source junction to remove holes from the body to be recombined in the source. In equilibrium, generation of holes must equal recombination of holes. Thus, the equilibrium OFF-state floating-body voltage ($V_{B,OFF}$) is greater than zero. $V_{B,OFF}$ forward biases the body-source junction diode enough to remove holes generated by thermal generation and impact ionization.

Figure 2-5 shows the SOI MOSFET turning ON. As the gate voltage rises, the excess gate-body depletion region holes are trapped in the body by the potential barrier of the body-source junction. The body voltage instantaneously rises to reflect the excess of holes in the body. This increased forward bias of the body-source diode begins to remove the excess holes from the body to be recombined in the source. As excess holes are removed over time, the body voltage tends to decrease. However, while holes are being removed through the body-source diode, TURN-ON-state impact ionization generation and thermal generation in depletion regions are adding holes to

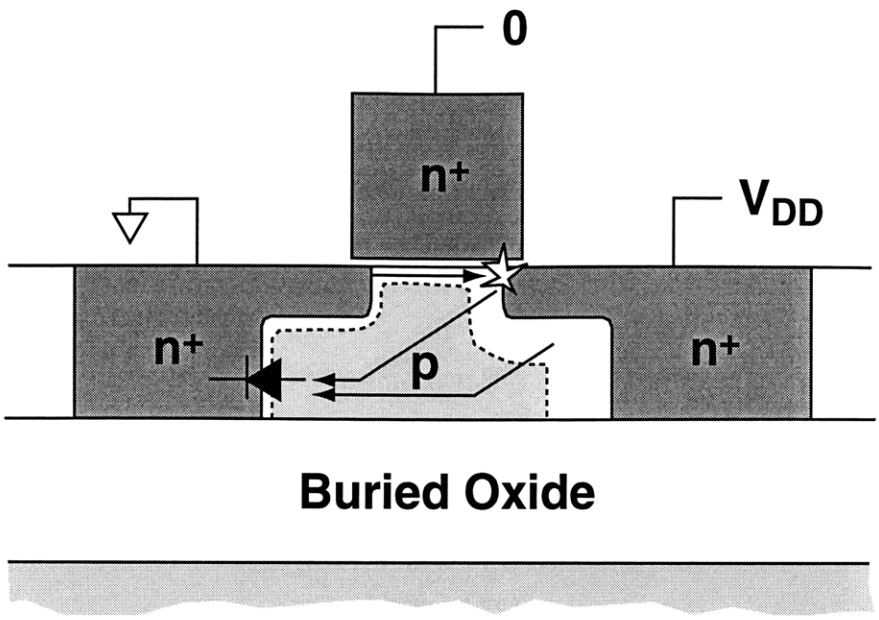


Figure 2-4. Schematic cross section of a SOI NMOSFET in the OFF-state. Arrows show impact ionization current and drain-body depletion region generation current flowing into the floating body to forward bias the body-source diode.

the body, tending to increase the body voltage. If the drain voltage stays at V_{DD}, as is done when measuring the DC gate transfer *I-V* characteristics, the body voltage will eventually reach an equilibrium DC value where the body-source diode removes the holes added the body by impact ionization and thermal generation. When the MOSFET is switching, however, the drain voltage actually falls so only the instantaneous value of drain current during TURN-ON switches the MOSFET from OFF to ON.

Figure 2-6 (a) shows this difference between the DC *I-V* characteristics and the instantaneous I-V characteristics. An example of the transient evolution from the instantaneous drain current value to the equilibrium DC value is shown in Figure 2-6 (b). Thus, due to the floating body, every time the terminal voltages are set to a bias point on an *I-V* characteristic, there will be an instantaneous value of drain current which over time changes and settles to an equilibrium DC value. The time-transient evolution of drain current from instantaneous to equilibrium DC values will be different for each bias point, since the instantaneous body voltage will be different at each bias point and the impact ionization and thermal generation rates will also be different at each bias point.

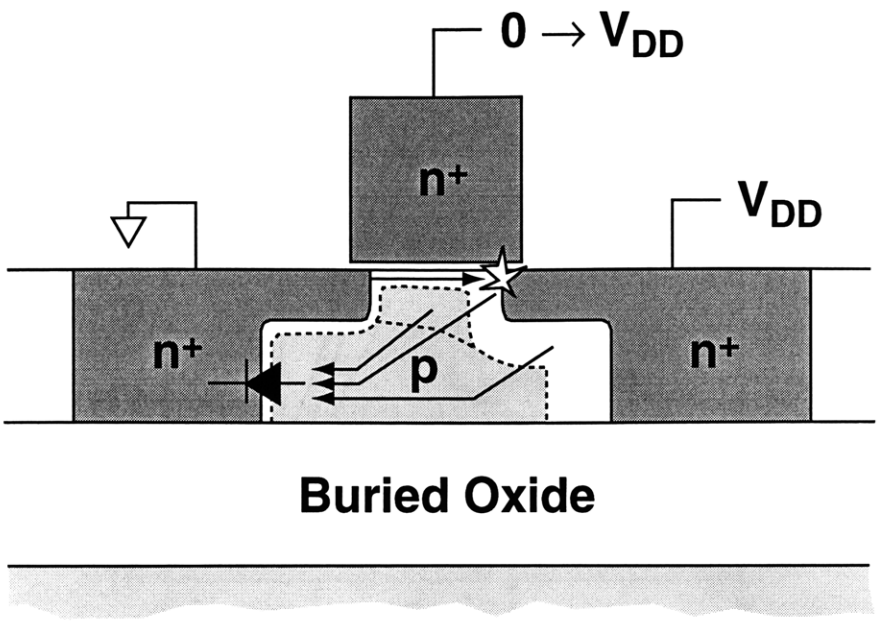
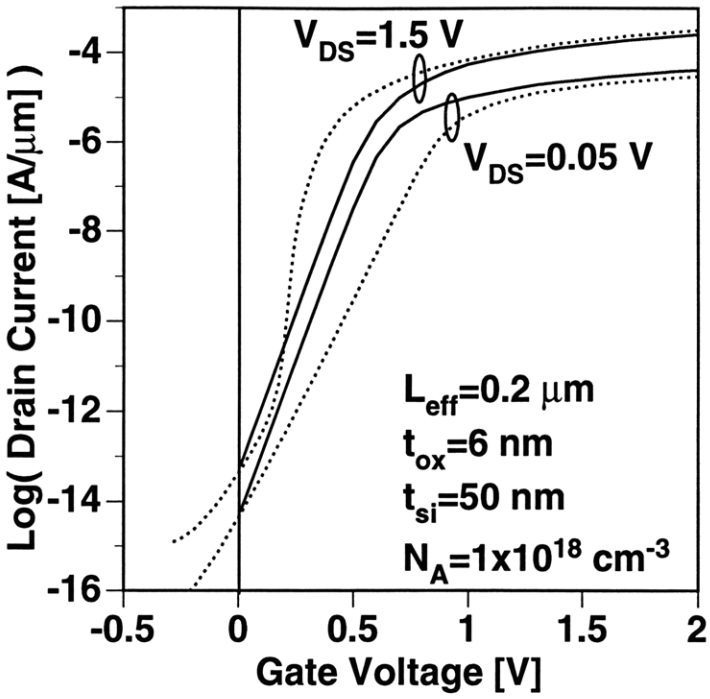


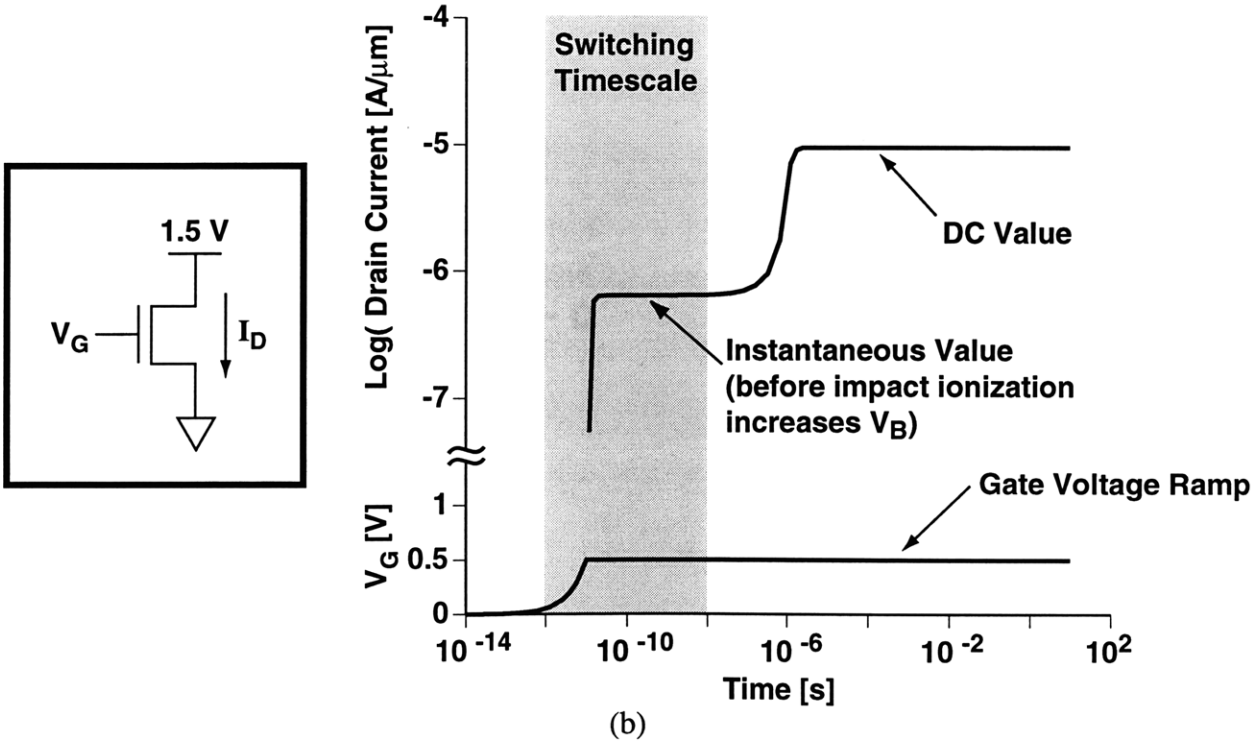
Figure 2-5. Schematic cross section of a SOI NMOSFET switching from the OFF-state to the ON-state. Arrows show gate-depleted body charge, impact ionization current, and drain-body depletion region generation current flowing into the floating body to forward bias the body-source diode.

This idea is made more clear by representing the SOI MOSFET floating body as an equivalent network as shown in Figure 2-7 (a). Rather than a body-contact connecting the body to the outside world, the floating body is in effect "connected" to the terminals by this network of capacitors, diodes, and current sources. The instantaneous AC value of the floating-body voltage is determined by the body-terminal and terminal-terminal capacitances shown in Figure 2-7 (b). The equilibrium DC body voltage and body hole content is determined by the network of body-source/drain diodes and dependent current sources representing impact ionization generation, gate-body space charge region thermal generation, and, if present, gate-body tunneling current, shown in Figure 2-7 (c).

From this representation of the floating body, three points are obvious. 1) Changing device design changes the capacitances, diodes, and generation sources. Thus, floating-body effects can significantly vary with device design. 2) Transient *I-t* characteristics are different depending on how the terminal voltages are switched. For example, the instantaneous body voltage obtained by fixing the drain voltage and pulsing the gate voltage will be different than fixing the gate voltage

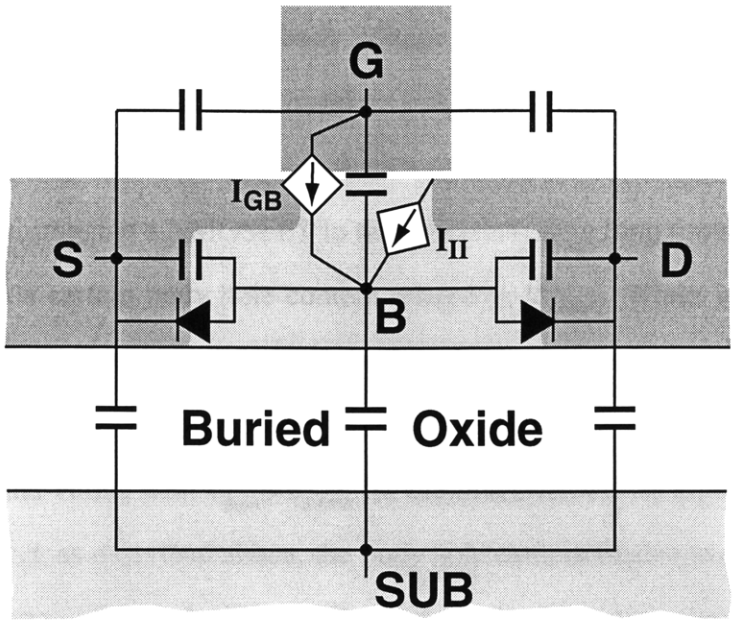


(a)

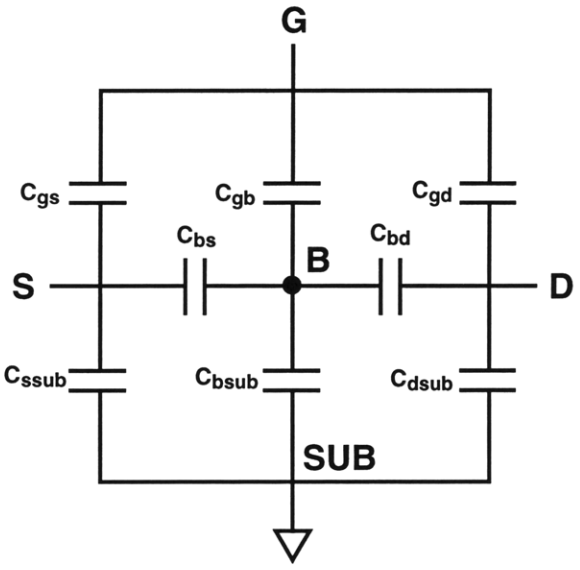


(b)

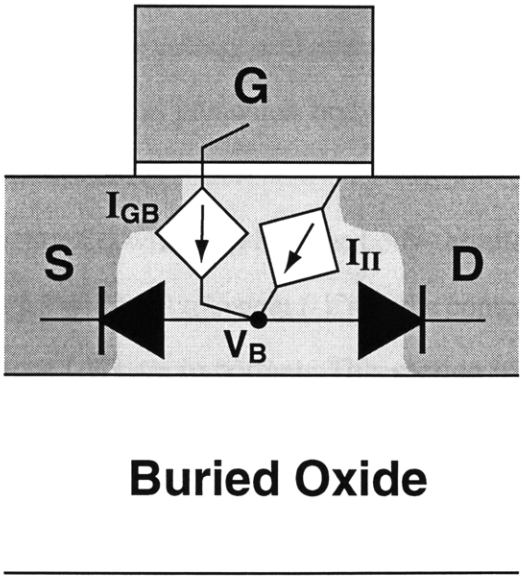
Figure 2-6. (a) Example of instantaneous transient (dashed lines) and DC (solid lines) gate subthreshold *I-V* characteristics for a partially-depleted SOI MOSFET. (b) Schematic showing a the transient drain current changing from the instantaneous value to the DC value for the bias point V_{DS}=1.5 V and V_{GS}=0.5 V in (a).



(a)



(b)



(c)

Figure 2-7. (a) Equivalent representation of the SOI MOSFET floating body. (b) AC network to determine the instantaneous floating-body voltage. (c) DC network to determine the steady-state floating-body voltage.

and pulsing the drain voltage. The DC body voltage will be the same for the two cases, but the instantaneous *I-V* characteristics and the transient behavior to reach DC will be different between the two cases. 3) There will be hysteresis in device operation.

For example, imagine a NMOSFET in the OFF-state for a long time having established an equilibrium. It has a certain body hole content related to $V_{B,OFF}$. While $V_{B,OFF}$ is a single value related to being in OFF-state equilibrium, V_B in the OFF-state can change with switching. Thus, when discussing switching, $V_{B,OFF}$ will be renamed the preswitch body voltage ($V_{B,pre}$) to indicate that it is not a singular value, with $V_{B,pre}=V_{B,OFF}$ for the first switch from equilibrium. Now, switch the MOSFET ON and, as described above, the body hole content begins to change during TURN-ON and in the ON-state. When the MOSFET returns to the OFF-state, the body hole content has changed, thus the preswitch body voltage has changed as well. The MOSFET will switch ON differently the next time due to a different preswitch body voltage and hole content. If it is left OFF for a long period of time, the body hole content and preswitch body voltage will return to $V_{B,OFF}$. If it is continually switched ON and OFF, the body hole content and preswitch body voltage will continually change until a "switching-steady-state" is established.

The following address these three points. First, section 2.2 will show how the body-terminal connections change with device design. Second, section 2.3 will treat transient *I-V*'s in the context of CMOS digital operation, in which only the instantaneous *I-V*'s are important. This section will then show how the body-terminal connections determine instantaneous *I-V*'s and how they vary with device design. And third, hysteretic switching behavior will be discussed in chapter 3.

2.2 Body-Terminal Connections

SOI MOSFET design variables include the gate length (L_G), gate oxide thickness (t_{ox}), silicon film thickness (t_{si}), buried oxide thickness (t_{box}), power supply voltage (V_{DD}), and body and source/drain doping profiles (N_A and N_D). Carrier lifetimes can also be a variable in SOI MOSFET's as they have been engineered with beneficial results. Temperature may be an addition design variable

because devices may operate at various temperatures across a chip. The following sections describe each type of body-terminal connection in Figure 2-7, and shows how they vary versus device design variables.

2.2.1 Terminal-Body Capacitances

As shown in Figure 2-7 (b), the instantaneous floating-body voltage is capacitively coupled to the gate, source, drain, and substrate terminal voltages. Out of the capacitance network, the gate-body, source-body, and drain-body capacitances are the most important and can be easily described. Terminal-terminal capacitances are parasitic capacitances which indirectly affect the body through the gate-body, source-body, drain-body, and substrate-body capacitances. Note that all the capacitances in Figure 2-7 (b) are non-recriprocal, e.g. gate-body does not equal body-gate capacitance [29]. The following describes the physics of each capacitance and how design variables and applied bias affect the relative size of the capacitance, per unit width.

The gate-body capacitance is made up of the gate oxide capacitance (C_{ox}) and the capacitance of the depletion region underneath the gate oxide (C_{depl}). The oxide capacitance increases with decreasing gate oxide thickness. The depletion capacitance increases with increasing body doping, due to a decrease in depletion layer thickness. Both capacitances increase with increased gate length. In the subthreshold region, the gate-body capacitance is the series combination of these two capacitances. At or above the threshold voltage, formation of the inversion layer tends to screen the body node from the gate node, thereby reducing the total value of the series combination.

The body-source/drain capacitances are made up of a capacitance associated with the junction depletion region, and a capacitance associated with diffusion of minority carriers through the junction (C_{diff}). The depletion capacitance increases with increased body and/or source/drain doping, due to a decrease in depletion layer thickness. The depletion capacitance also increases with increased silicon film thickness. The diffusion capacitance can be though of as the small signal response of majority carriers to injection of minority carriers. Because it is directly related to the injection of

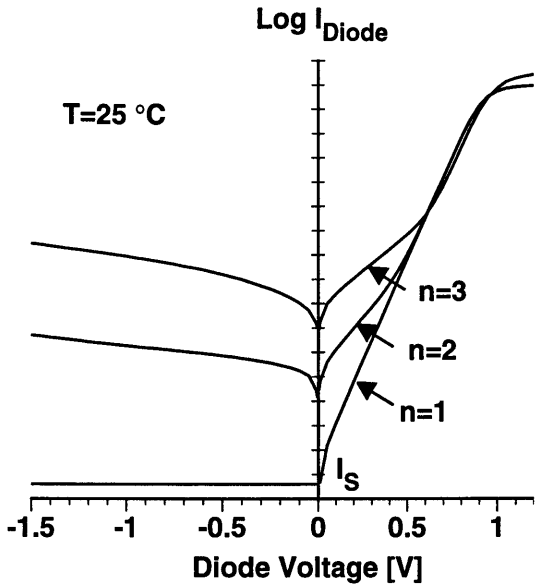


Figure 2-8. Schematic of diode *I-V* characteristics representing an ideal diode (n=1), a nonideal diode with generation-recombination current (n=2), and a nonideal diode with tunneling current (n=3). Each vertical division represents a factor of ten.

minority carriers, it increases with decreasing body doping (assuming the source doping fixed and much greater than the body doping) and increases exponentially by the factor qV_B/kT, where q is coulombic charge, V_B *is* the body voltage, k is Boltzmann's constant, and T is the temperature. At small forward bias the diffusion capacitance is typically smaller than the depletion capacitance. However, at high forward bias the diffusion capacitance can be much greater than the depletion capacitance [30].

The depletion capacitances described above have a weak temperature dependence, increasing slightly with increasing temperature. The diffusion capacitance, however, has strong temperature dependence and increases with temperature because injected minority carriers (diode current) increases with temperature. None of the capacitances vary with carrier lifetime.

2.2.2 Body-Source/Drain Diodes

Body-source/drain diode *I-V* characteristics are greatly affected by channel and source/drain doping, operating temperature, and carrier lifetime. Both the forward and reverse diode

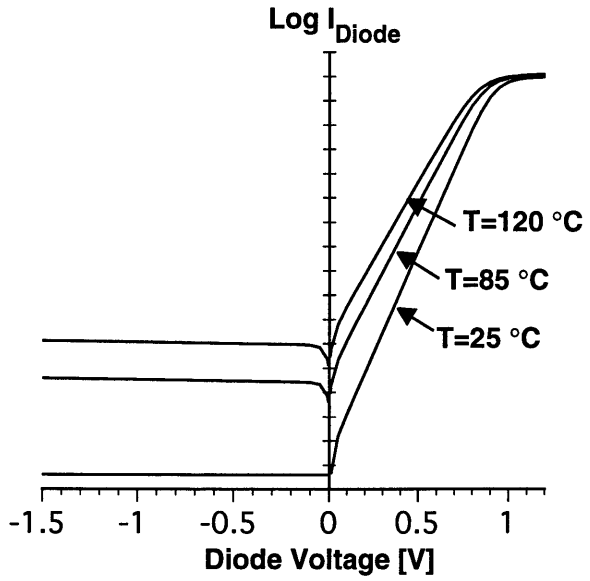


Figure 2-9. Diode *I-V* characteristics of the n=1 diode in figure 2-8, show at T=25, 85, and 120 °C. Each vertical division represents a factor of ten.

characteristics are important in SOI MOSFET operation.

Forward and reverse diode currents are made up of three components, the ideal diode current, Shockley-Read-Hall (SRH) current which accounts for thermal generation-recombination (*R-G*) in the junction depletion region, and a tunneling current component. This is shown schematically in Figure 2-8, which shows *I-V*'s of diode with only ideal diode current (n=1), a diode with ideal diode current and SRH current (n=2), and a diode with ideal diode current, SRH current, and tunneling current (n=3). n is the "ideality factor" of the diode, and is related to the diode current components as follows.

Ideal diode current is characterized by forward current increasing exponentially with applied forward voltage in the form $I_{DIODE}=I_S[exp(qV_A/nkT)-1]$, where I_S is the saturation current which depends on doping, and n=1 for ideal. For negative applied bias, the ideal diode reverse current is approximately the saturation current, $-I_S$. Assuming the source/drain doping is fixed and greater than the body doping, I_S decreases for increasing body doping. I_S increases with increasing temperature as shown in Figure 2-9, but is not dependent of carrier lifetime.

Forward bias SRH current is made up of recombination current due to an excess of carriers

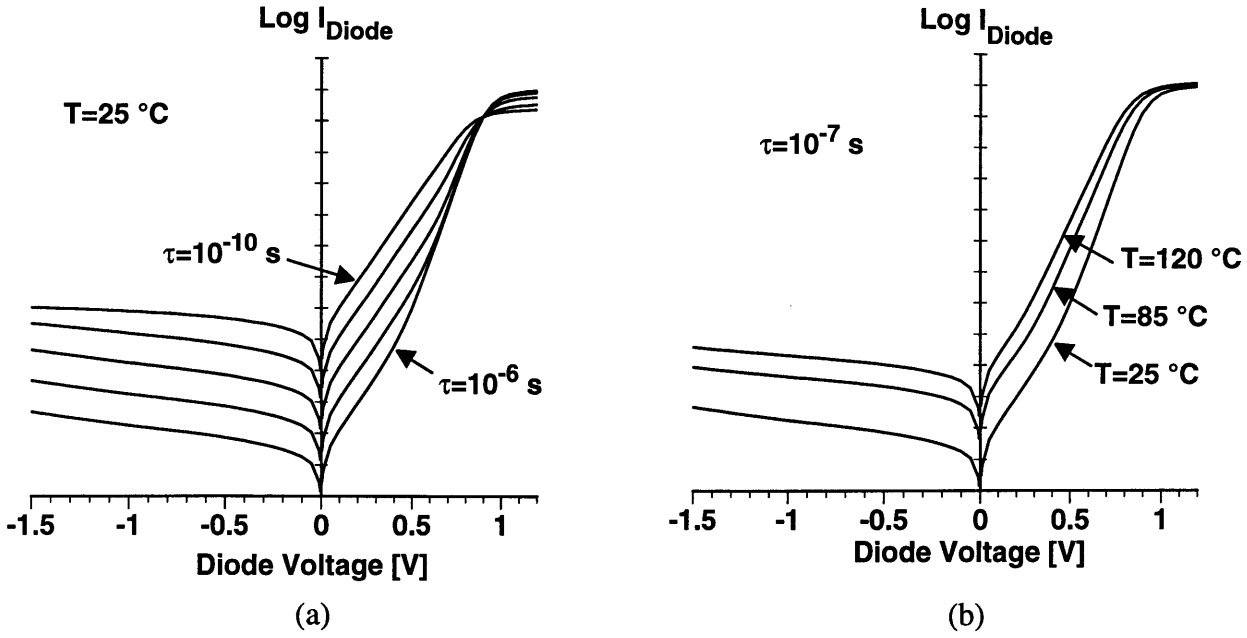


Figure 2-10. Diode *I-V* characteristics of the n=2 diode in figure 2-8. Shown in (a) at fixed temperature with $\tau=10^{-10}$, 10^{-9}, 10^{-8}, 10^{-7}, and 10^{-6} seconds, and in (b) with fixed carrier lifetime at T=25, 85, and 120 °C. Each vertical division represents a factor of ten.

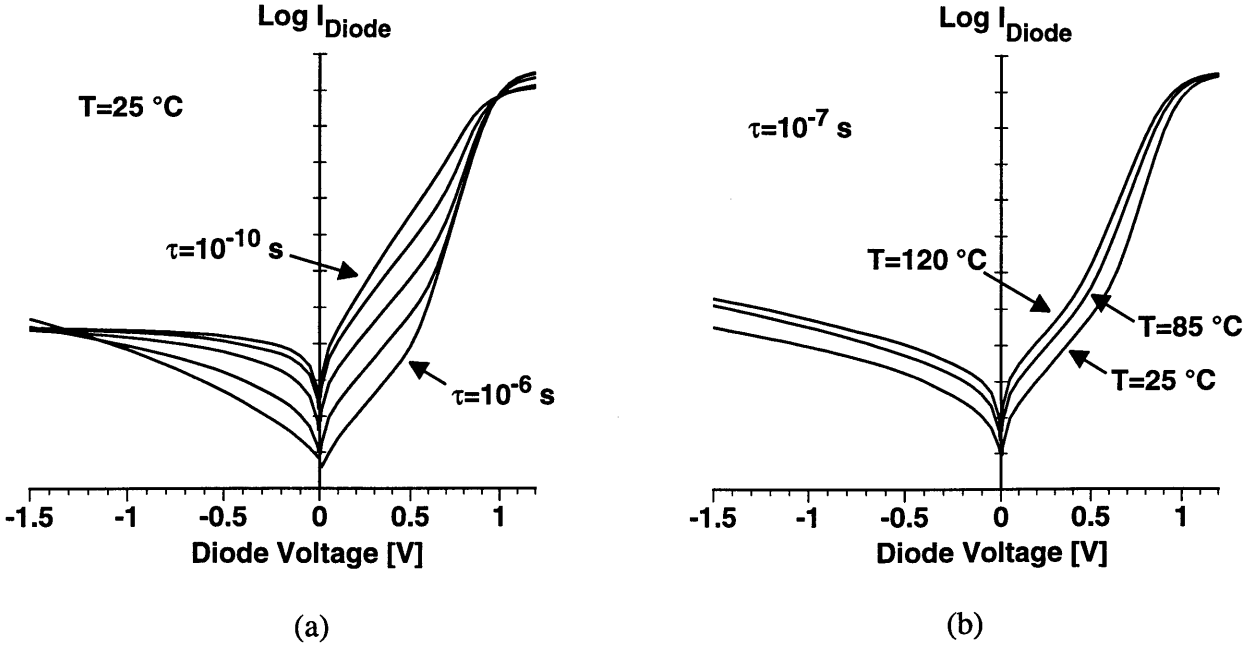


Figure 2-11. Diode *I-V* characteristics of the n=3 diode in figure 2-8. Shown in (a) at fixed temperature with $\tau=10^{-10}$, 10^{-9}, 10^{-8}, 10^{-7}, and 10^{-6} seconds, and in (b) with fixed carrier lifetime at T=25, 85, and 120 °C. Each vertical division represents a factor of ten.

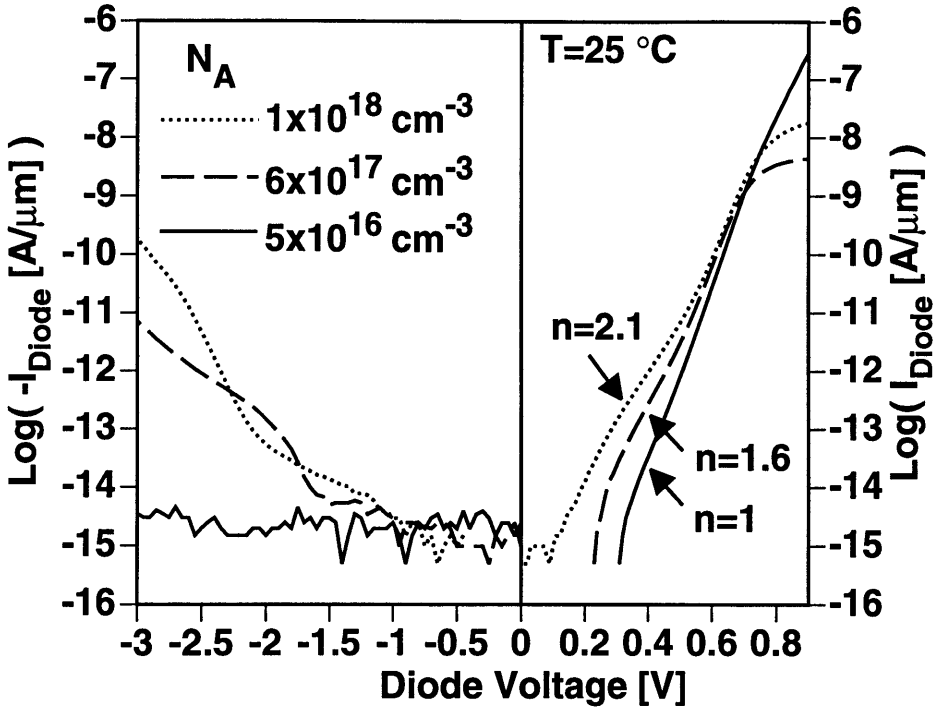


Figure 2-12. Measured body-source/drain diode *I-V* characteristics. Diode current is measured in body-contacted SOI MOSFET's with V_G=-1 V. Three devices with three different body doping levels are shown. Higher doping levels exhibit more nonideal behavior resulting in higher ideality factor n. The N_A=1x10^{18} and 6x10^{17} cm^{-3} have t_{si}=66 nm, and the N_A=5x10^{16} cm^{-3} devices has t_{si}=115 nm.

in the forward biased depletion region. It is of the form $I_{DIODE,recomb}=I_R[exp(qV_A/nkT)]$, where I_R is the recombination saturation current. Because forward SRH current depends on carrier recombination, I_R and n depend the location of *R-G* centers in the bandgap, electron and hole capture cross sections, and doping. Due to these variables, $1<n\leq 2$. n=2 assuming the R-G centers are at midgap and equal hole and electron capture cross sections. Reverse bias SRH current is made up of generation current due to deficit of carriers in the reverse biased depletion region. Reverse bias SRH current is $-I_R$ at zero bias, and increases with increasing reverse bias due to increased depletion width.

Since SRH currents result from thermal generation-recombination processes, they are greatly affected by carrier distributions, carrier lifetime, and temperature. Assuming the source/drain doping is fixed and greater than the body doping, I_R increases with increasing body doping because this increases the area in the depletion region where the hole concentration is nearly equal to the electron

concentration ($p \approx n$), which maximizes recombination [30]. Carrier lifetime and temperature also determines the value of I_R. The n=2 diode from Figure 2-8 is shown at five different carrier lifetimes in Figure 2-10 (a), and at three temperatures in Figure 2-10 (b). For a fixed carrier lifetime, however, forward ideal diode current increases faster than forward SRH current, as shown in Figure 2-10 (b).

Tunneling current occurs at high doping levels. At high doping levels, the depletion region width is sufficiently thin such that under zero bias and forward bias conditions, electrons can tunnel from the conduction band to traps in the depletion region, and recombine with holes. This forward bias trap-assisted recombination current is characterized by forward diode current ideality factor n>2 [31]. Reverse tunneling current (Zener breakdown due to band-band tunneling) also occurs at high doping levels where the depletion region is sufficiently narrow for an electron in the valence band to tunnel to the conduction band on the other side, thus creating electron-hole pairs which make up reverse bias generation current. Tunneling current increases with reverse bias despite an increase in depletion width because the increased electric field decreases the tunneling barrier [32].

Assuming the source/drain doping is fixed at 10^{20} cm^{-3} and the junction is fairly abrupt, body doping levels of 10^{18} cm^{-3} begin to show tunneling characteristics. Above $2x10^{18}$ cm^{-3}, the tunneling current component begins to increase rapidly [33], to the point where at high enough body doping levels the forward and reverse diode *I-V* characteristics may be nearly symmetric at low current values. This is shown in Figure 2-11 (a), which shows the diode with n=3 in Figure 2-8 at five values of carrier lifetime. At the longest carrier lifetime, SRH current is suppressed and the forward and reverse diode characteristics are nearly symmetric at low current values. At shorter carrier lifetimes, the SRH current becomes much greater than the tunneling current component. Figure 2-11 (b) shows the temperature dependence of the tunneling current. For carrier lifetimes long enough for tunneling current to be greater than SRH current, the tunneling current component retains its characteristic shape as the temperature is increased.

The above diode *I-V*'s in Figures 2-8 to 2-11 were created in the 2-D numerical simulator MEDICI with nearly abrupt junctions representative of modern MOSFET body-source/drain junctions. The trends described above are observed in measured body-source/drain diodes as shown

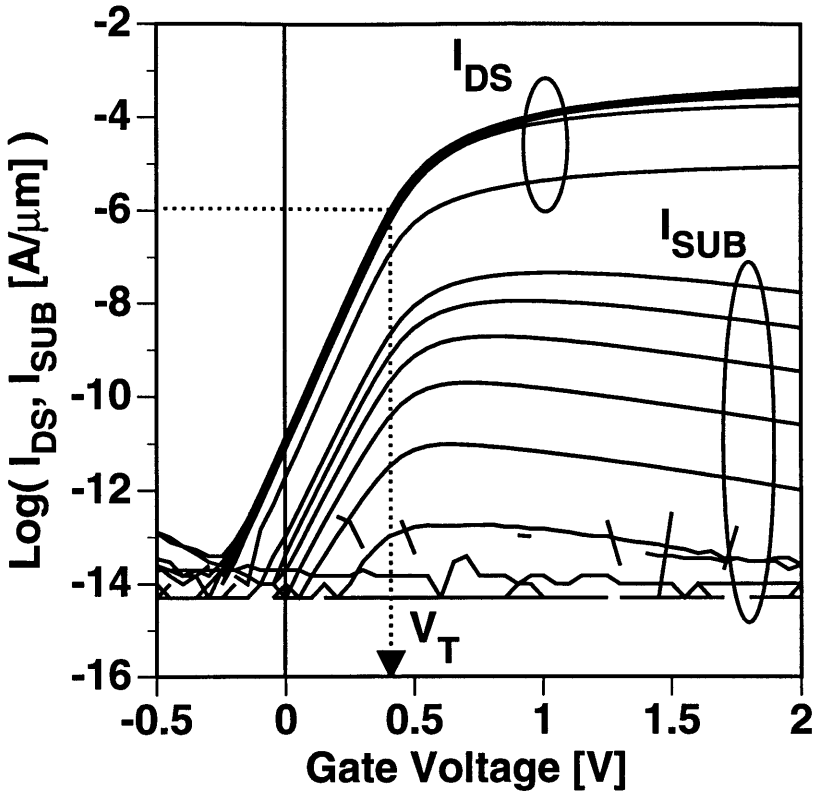


Figure 2-13. Measured gate transfer *I-V* characteristics and substrate current due to impact ionization generation. I_{DS}-V_{GS} and I_{SUB}-V_{GS} curves correspond from top to bottom to V_D=2.01, 1.81, 1.61, 1.41, 1.21, 1.01, 0.81, and 0.01 V. The measured device is a bulk MOSFET with L_{eff}=0.1 μm. The saturation threshold voltage defined at a constant current of $10^{-7}/L_{eff}$ A/μm is shown.

in Figure 2-12. These diode *I-V* characteristics were measured from the body to the source/drain in body-contacted SOI NMOSFET's with substrate ground and the gate biased at -1 V.

And one final note: due to ion implantation, body doping levels can be highly nonuniform. Carrier lifetimes can vary along the junction by varying the tilt angle of a lifetime-killing ion implant [34]. The body-source/drain diode behavior reflects the physical makeup of the junction, i.e. the tunneling current component will flow near the peak of the body doping, the SRH current component will flow mostly in the area of the junction with shortest carrier lifetime, and the ideal component will be highest in the lowest doped body regions of the junction. This is an important point to remember when looking at how body-source/drain diode currents scale with silicon film thickness.

2.2.3 Impact Ionization Current Source

In the high electric field of the reverse biased body-drain junction, carriers can gain enough energy to cause ionization of atoms upon impact with that atom, creating an electron-hole pair. This process is characterized by an impact ionization rate. This rate multiplied by the drain current determines how many electron-hole pairs are generated. Thus the TURN-ON-state impact ionization current is much higher than the OFF-state impact ionization current. The impact ionization rate changes with drain-body voltage, gate-body voltage, and body-source/drain doping profile since these variables determine the peak electric field in the drain-body depletion region.

Figure 2-13 shows impact ionization current measured in a bulk MOSFET versus gate voltage for various drain voltages. Again, the total impact ionization current is the product of the impact ionization rate and the drain current. The peak impact ionization rate occurs at zero gate voltage, but there is little subthreshold drain current. As the gate voltage increases, the peak electric field at the surface decreases as the vertical field from the gate to body affects the tangential field from the drain to body. However, drain current is increasing as the gate voltage is increasing. Thus, the peak impact ionization current occurs at the onset of strong inversion slightly above the threshold voltage as shown in Figure 2-13 [35].

Also keep in mind that impact ionization current can vary with body voltage as well. Increased body voltage decreases the reverse bias of the drain-body junction thereby decreasing the peak electric field and impact ionization rate. However, drain current increases with increased body voltage thereby increasing impact ionization current.

In general, the impact ionization rate increases with increased body doping since this makes the electric field higher. The impact ionization rate is a weak function of temperature, but the impact ionization current decreases with increasing temperature because the drain current decreases due to mobility degradation. The magnitude of impact ionization current is not affected by changes in carrier lifetime.

2.2.4 Gate-Body Current Source

The body-gate current source represents SRH generation of electron-hole pairs in the depletion region under the gate. It also represents direct tunneling of carriers through the gate oxide, if the gate oxide is sufficiently thin (less than 2 nm) [32].

Without direct tunneling of carriers through the gate oxide, the gate-body current source is much smaller than the reverse bias diode SRH generation current. This is because the depletion region under the gate does not have a region where p≈n, which maximizes *R-G* statistics. With direct tunneling of carriers through the gate oxide, the gate-body current source can be significant because the gate-polysilicon-oxide-body can display diode-like behavior.

2.3 Floating-Body Effects in CMOS Digital Operation

The above description of the floating body suggests an endless list of possible transient floating-body effects. However, the focus of this thesis is how the floating body affects the design of SOI CMOS technologies for logic applications. The following begins this assessment by treating SOI MOSFET *I-V* characteristics in the context of CMOS digital operation, where only the instantaneous *I-V*'s are important.

2.3.1 CMOS Digital Operation

CMOS digital circuits are made up of combinations of pull-up networks, pull-down networks, and pass-gates. For given logic inputs, these networks either pull-up or pull-down the voltage of the output node. To illustrate MOSFET digital operation, the following analysis will be based on the simplest pull-down network: a NMOSFET in a CMOS inverter. Figure 2-14 shows a simplified schematic of the CMOS inverter switching from OUT-HI to OUT-LO. This occurs because the input gate voltage of the NMOSFET rises, the drain current rises, discharging the load capacitor (C_L) at the output node causing the drain voltage to fall. The timescale of this gate-rise and drain-

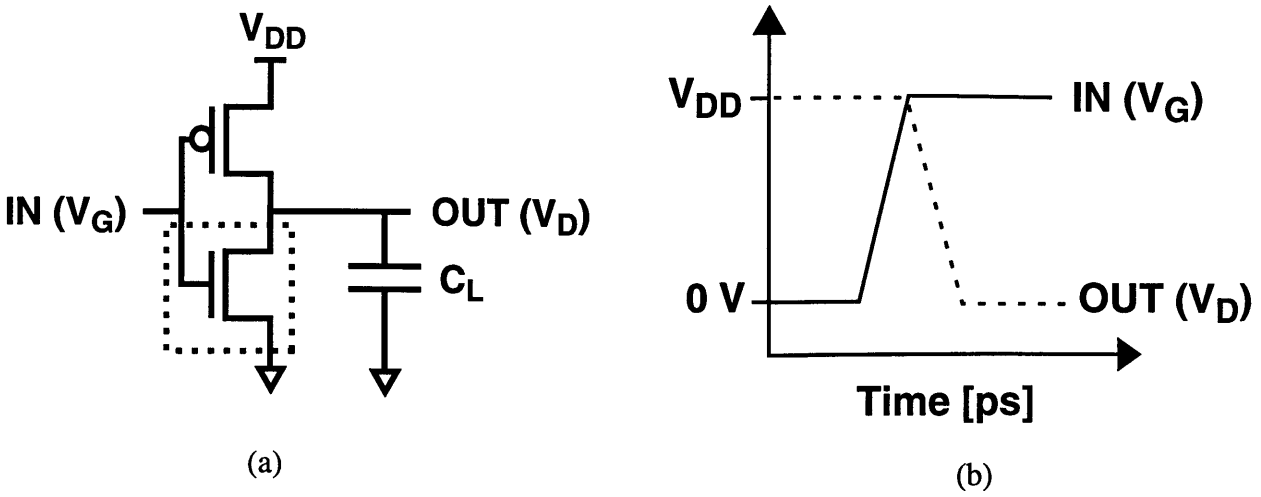
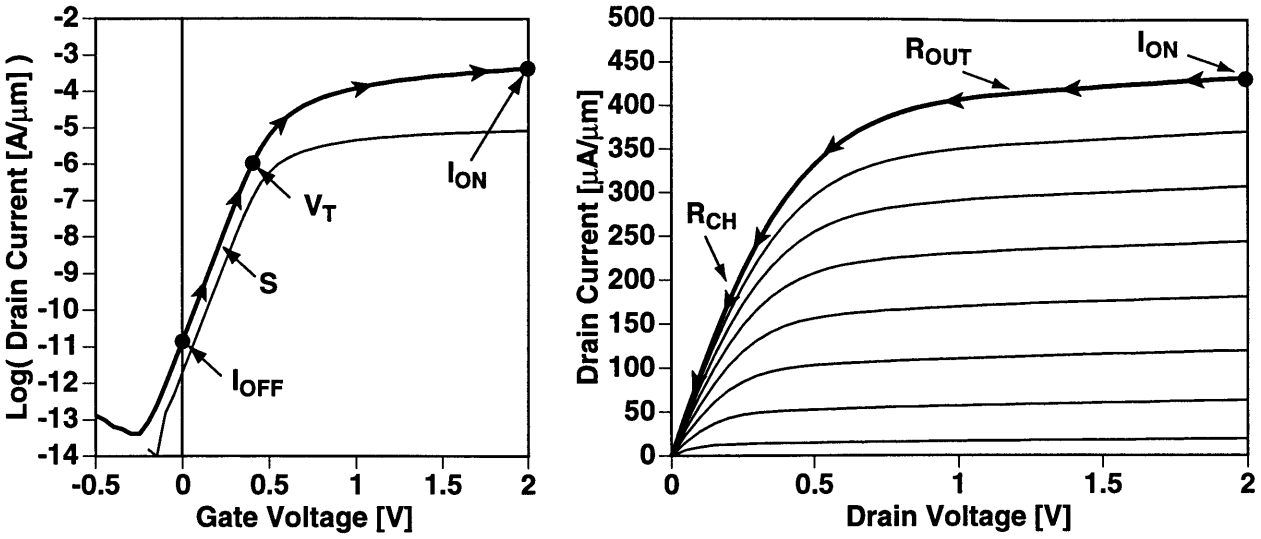


Figure 2-14. (a) Schematic showing the NMOSFET in a CMOS inverter. (b) Schematic showing the switching waveforms on the NMOSFET during a CMOS inverter switch from OUT-HI to OUT-LO which turns the NMOSFET from OFF to ON.

fall is very short. For example, one half-period of a 1 GHz clock is 500 ps. In high performance designs, an order of ten logic transitions can be expected to occur in this half-period. This results in logic rise and fall transitions on the order of tens to hundreds of picoseconds, depending on output load. Thus, only the instantaneous *I-V*'s are important for SOI MOSFET's when considering a single switching event (multiple switching events will be discussed in chapter 3).

To illustrate the relationship of CMOS switching to device *I-V* characteristics, the switching event in Figure 2-14 is schematically superimposed on the MOSFET *I-V* characteristics in Figure 2-15. Figure 2-15 (a) shows the gate transfer *I-V* characteristics, with the I_D-V_G locus illustrating the rise in gate voltage shown in bold. Figure 2-15 (b) shows the output characteristics, with the I_D-V_D locus illustrating the fall of the drain voltage shown in bold. Highlighted in Figure 2-15 (a) are the important MOSFET parameters in digital logic. They are the OFF-state leakage current (I_{OFF}), the inverse subthreshold slope (S), which together determine the threshold voltage (V_T), which in turn determines the maximum drain current during TURN-ON (I_{ON}) and thus switching speed. I_{ON}, the output resistance (R_{OUT}), and the channel resistance (R_{CH}) are the important MOSFET parameters which determine the speed at which the drain voltage can fall. In digital logic, the speed at which



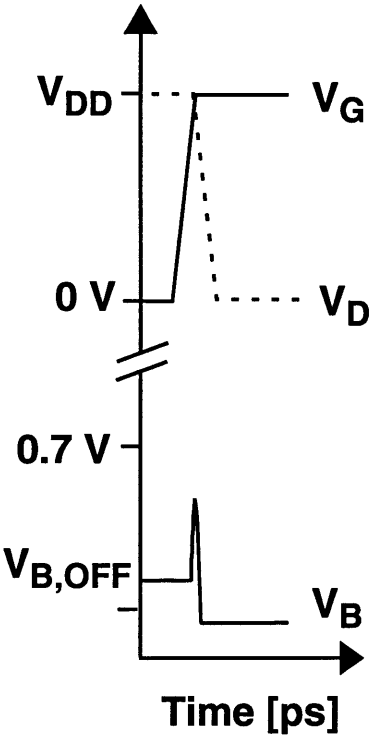
(a) (b)

Figure 2-15. (a) Schematic NMOSFET gate transfer *I-V* characteristic and the corresponding I_D-V_G locus corresponding to figure 2-14. (b) Schematic NMOSFET output *I-V* characteristic and the corresponding I_D-V_D locus corresponding to figure 2-14.

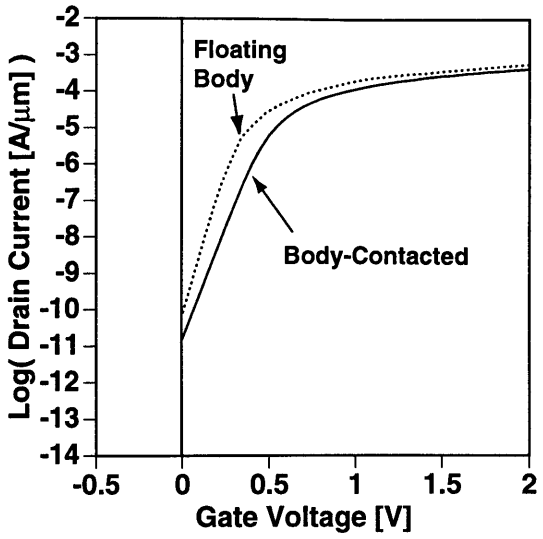
the drain voltage can fall determines the rate of change of the output node voltage, which is connected to an input of the next logic stage.

A well designed MOSFET should possess minimum I_{OFF}, maximum I_{ON}, minimum V_T and S, minimum R_{CH} and maximum R_{OUT}. Minimizing I_{OFF} reduces standby power consumption while minimizing V_T and S increases I_{ON} and the speed at which the MOSFET turns ON. Maximizing I_{ON} and R_{OUT} and minimizing R_{CH} maximizes the speed of the fall of the drain voltage, and thus maximizes the speed of switching the next stage.

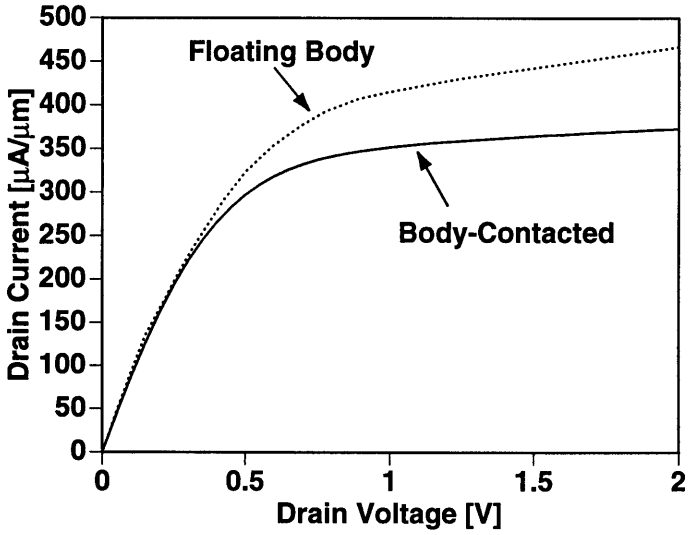
Due to the short switching timescale, it is the I_{OFF}, S, V_T, I_{ON}, R_{OUT} and R_{CH} from the instantaneous *I-V*'s which are of interest in SOI MOSFET's. Any change in body hole content will be assumed negligible (non-negligible changes in body hole content over many switches will be discussed in chapter 3). Figure 2-16 contrasts how the floating body SOI MOSFET reacts to switching compared to a bulk or body-contacted MOSFET also shows the floating-body. Figure 2-16 (a) shows the switching waveforms applied at the gate and drain. Also shown is the floating-body voltage, which starts off at $V_{B,OFF}$. Due to $V_{B,OFF}$>0 V, I_{OFF} is higher when the body is floating



(a)



(b)



(c)

Figure 2-16. (a) Schematic showing the switching waveforms on a SOI NMOSFET during a CMOS inverter switch from OUT-HI to OUT-LO and corresponding floating-body voltage. (b) Sample gate transfer characteristic showing the difference between floating-body and body-contacted operation. (c) Sample drain output characteristic showing the difference between floating-body and body-contacted operation.

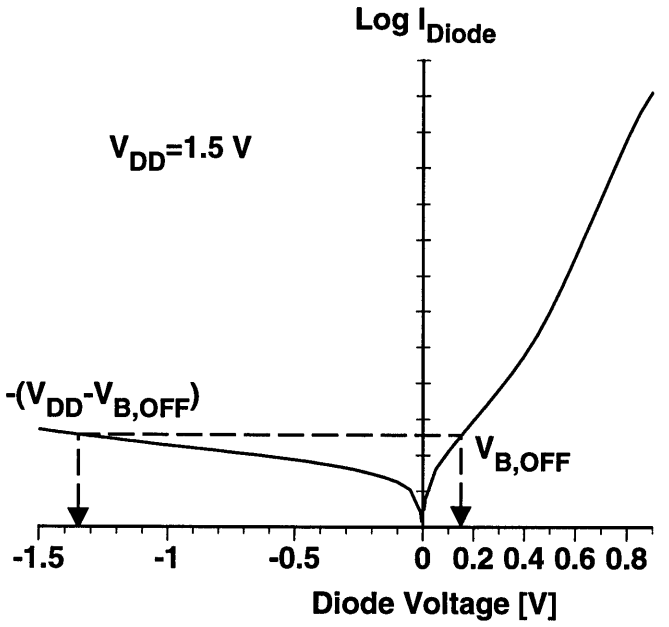
as shown in Figure 2-16 (b). The body voltage is capacitively pulled up by the gate as the gate voltage increases. This tends to decrease S, decrease V_T, and increase I_{ON}. When the drain falls, however, the body voltage is capacitively pulled down which decreases R_{OUT} compared to the body-contacted case. As will be shown in the following sections, the diode and impact ionization characteristics determine $V_{B,OFF}$ and I_{OFF} while the magnitude of capacitive coupling determines S, V_T, I_{ON}, and R_{OUT}. The following sections discuss how device design changes these parameters.

2.3.2 OFF-Current

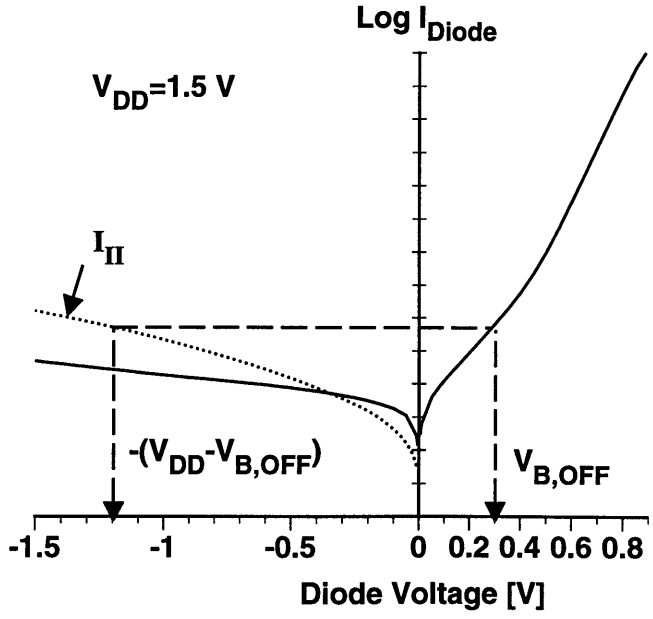
As shown in Figure 2-15 (a), I_{OFF} is the subthreshold current in the OFF-state at V_G=0 V and V_D=V_{DD}. MOSFET subthreshold drain current is strong function of the applied body-source voltage, V_{BS}. Thus, in the OFF-state, I_{OFF} is a strong function of V_{BS}=$V_{B,OFF}$. In general, I_{OFF} increases with increased $V_{B,OFF}$, and decreases with increased N_A.

In the SOI MOSFET, $V_{B,OFF}$ is set by the back-to-back body-source/drain diodes and impact ionization and gate-body currents. Figure 2-17 shows how $V_{B,OFF}$ is determined graphically. If impact ionization and gate-body currents are not significant, $V_{B,OFF}$ is simply the body voltage which makes the current through the body-source diode biased at $V_{B,OFF}$ equal to the current through the body-drain diode reverse biased at (V_{DD}-$V_{B,OFF}$), as shown in Figure 2-17 (a). If impact ionization and/or gate currents are significant, they can be added to the reverse bias diode characteristics and $V_{B,OFF}$ can be determined similarly as shown in Figure 2-17 (b). Note that significant impact ionization increases $V_{B,OFF}$. From Figure 2-17, it is also clear that decreased V_{DD} reduces $V_{B,OFF}$ due to reduction of voltage to drop across the back-to-back diodes and reduction of impact ionization current.

From Figure 2-17, it is also clear that the current levels of the diodes and the symmetry of the forward to reverse diodes are also important in determining $V_{B,OFF}$. Higher currents flowing through the body-source/drain diodes tend to make impact ionization current less significant, thereby reducing $V_{B,OFF}$. As shown in Figures 2-9 through 2-11, higher diode current levels are achieved at higher temperature (which also decreases impact ionization current) and shorter carrier lifetimes.



(a)



(b)

Figure 2-17. Graphical representation of the OFF-state floating body voltage without significant impact ionization (a), and with significant impact ionization (b). Note the increase in $V_{B,OFF}$ with significant impact ionization.

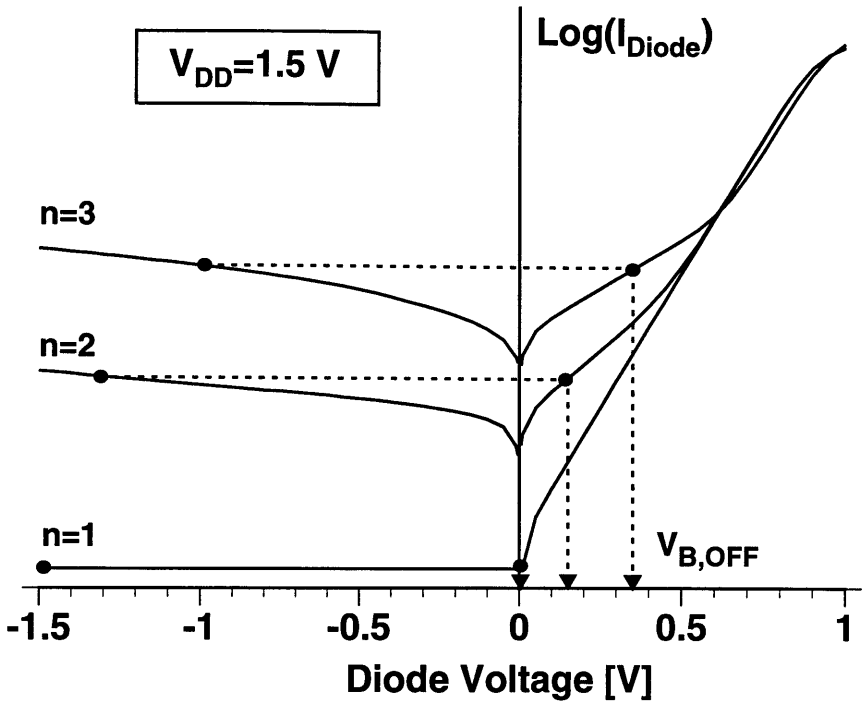


Figure 2-18. Schematic showing the dependence of $V_{B,OFF}$ on the symmetry between forward and reverse bias diode characteristics. n refers to forward bias diode ideality factor.

Increasing body doping also raises diode current with tunneling current, but increased symmetry between forward and reverse diode I-V characteristics. Figure 2-18 shows that increasing diode symmetry increases $V_{B,OFF}$. This makes SOI MOSFET's require more body doping to control I_{OFF}, because while I_{OFF} tends to decrease with increasing body doping, diode symmetry increases thereby increasing $V_{B,OFF}$ and I_{OFF}. But as shown in Figure 2-11 (b), diode symmetry can be made more independent of body doping by decreasing the carrier lifetime.

2.3.3 Gate Transfer and Output Characteristics

The subthreshold slope, S, is defined as the slope of subthreshold current versus gate voltage in a semilog gate transfer I-V plot. S is in the units of inverse slope, millivolts per decade of subthreshold current (mV/dec). For the same I_{OFF}, steepening the subthreshold slope (minimizing S) decreases V_T thereby increasing I_{ON}.

Figure 2-19 compares the capacitance networks for bulk and SOI MOSFET's which are

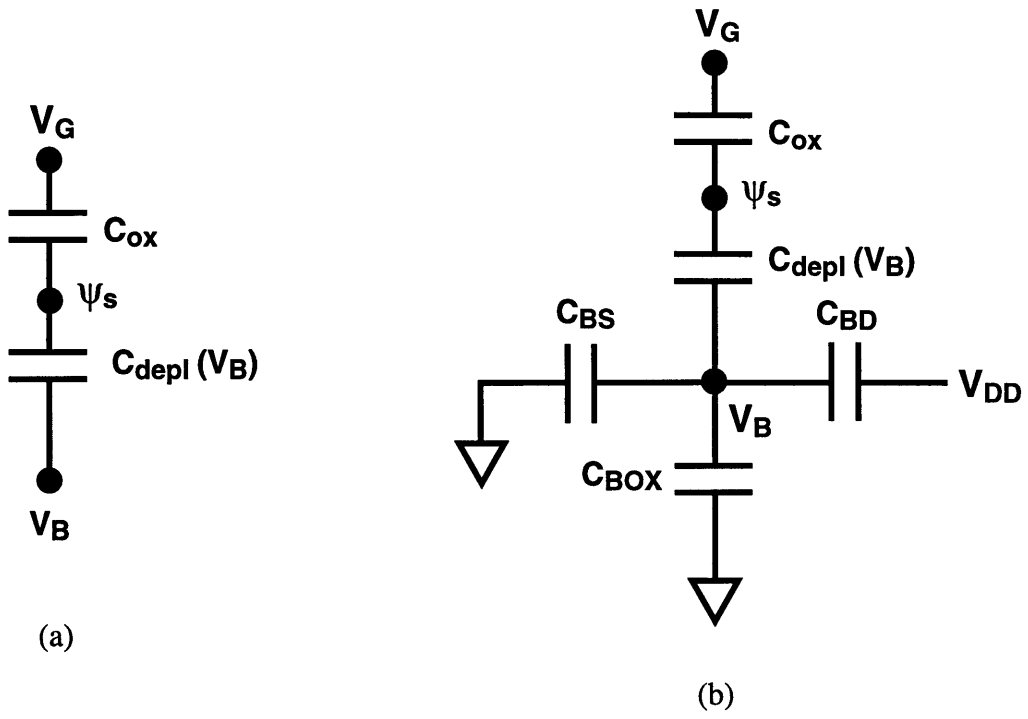


Figure 2-19. Relevant capacitive divider for determination of subthreshold slope for a bulk or body-contacted SOI MOSFET (a), and for a floating-body SOI MOSFET (b).

relevant for determination of subthreshold current. Recall that the subthreshold current is due to diffusion of electrons across the channel from source to drain. The number of available electrons depends on the surface potential, ψ_s, which can be modulated by the gate voltage resulting in the expression for subthreshold current shown in equation 2-3. Upon first inspection, SOI MOSFET's should always have steeper slope than a bulk MOSFET starting at the same $V_B=V_{B,OFF}$ because the floating-body voltage is capacitively coupled to the gate voltage. As the gate voltage increases, the body voltage increases thereby increasing the subthreshold current. This results in a steeper subthreshold slope.

However, SOI MOSFET's do not necessarily have steeper subthreshold slope compared to bulk MOSFET's with V_B set to zero by the body contact. This is because $V_B>0$ degrades subthreshold slope (increases S) by increasing C_{depl}, thereby decreasing ψ_s. SOI MOSFET gate-body capacitive coupling is only guaranteed to enhance the subthreshold slope compared to a bulk MOSFET with

the same V_B=$V_{B,OFF}$ applied to the body contact. At $V_{B,OFF}$, the subthreshold slope depends on the values of the capacitances in the network shown in Figure 2-19 (b). In particular, the relative magnitude of the gate-body capacitance versus the body-source capacitance, along with $V_{B,OFF}$, mainly determine S. This is because the floating-body voltage above $V_{B,OFF}$ is mainly determined by the capacitive divider between the gate-body capacitance and the body-source capacitance.

Given an I_{OFF} and S, the threshold voltage, V_T, can be determined. V_T delineates the transition from moderate to strong inversion current. It is convenient to define V_T for all devices with a constant current level near the transition from subthreshold current to inversion current. The constant current used in this thesis will be of ($10^{-7}/L$) A per micron of width.

As shown in Figure 2-15 (b), the output characteristics should reflect I-V's during the fall of the drain voltage immediately after the rise in gate voltage. Recall from Figure 2-16 that the rising gate voltage has the effect of pulling up the body voltage. This tends to steepen the subthreshold slope, lowering the threshold voltage, and increasing I_{ON} compared to a bulk MOSFET. This is a desirable effect since this increases switching speed. As the drain voltage falls, however, the body voltage is pulled down. This has effectively increases the threshold voltage as the drain voltage falls. This results in a steeper output characteristic, decreasing R_{OUT} and increasing R_{CH} compared to a bulk MOSFET. This is undesirable because it reduces the rate at which the drain voltage can fall thereby reducing the speed of the next stage.

The effect of the floating body on SOI MOSFET switching can be summarized in Figures 2-20 and 2-21. For a given gate length, gate oxide thickness, and supply voltage, only $V_{B,OFF}$ and the gate-body versus body-source/drain capacitance ratio can be varied. Assuming the same gate-body versus body-source/drain capacitance ratio, Figure 2-20 shows that decreased $V_{B,OFF}$ decreases I_{OFF} and decreases S. Assuming the same $V_{B,OFF}$, Figure 2-21 shows that changing the body-source/drain capacitance can greatly affect the I-V characteristics. Figure 2-21 (a) compares the body voltage through on switch cycle as defined in Figures 2-14 and 2-16. For lower body-source/drain capacitance, the body voltage can be pulled higher when the gate rises and is pulled down less

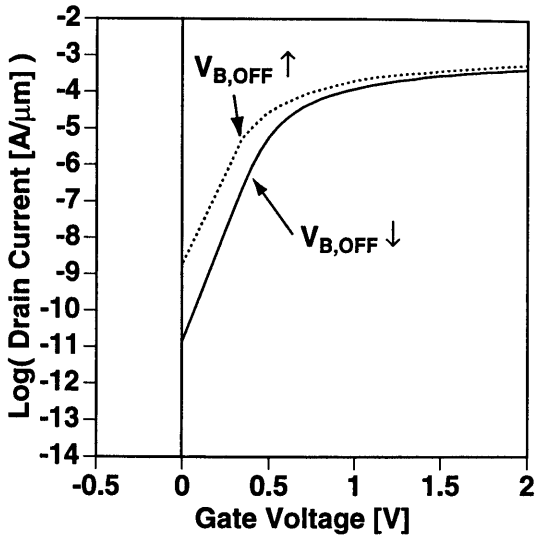


Figure 2-20. Schematic showing gate transfer *I-V* characteristic changing versus $V_{B,OFF}$ independent of changes in body-source/drain capacitance.

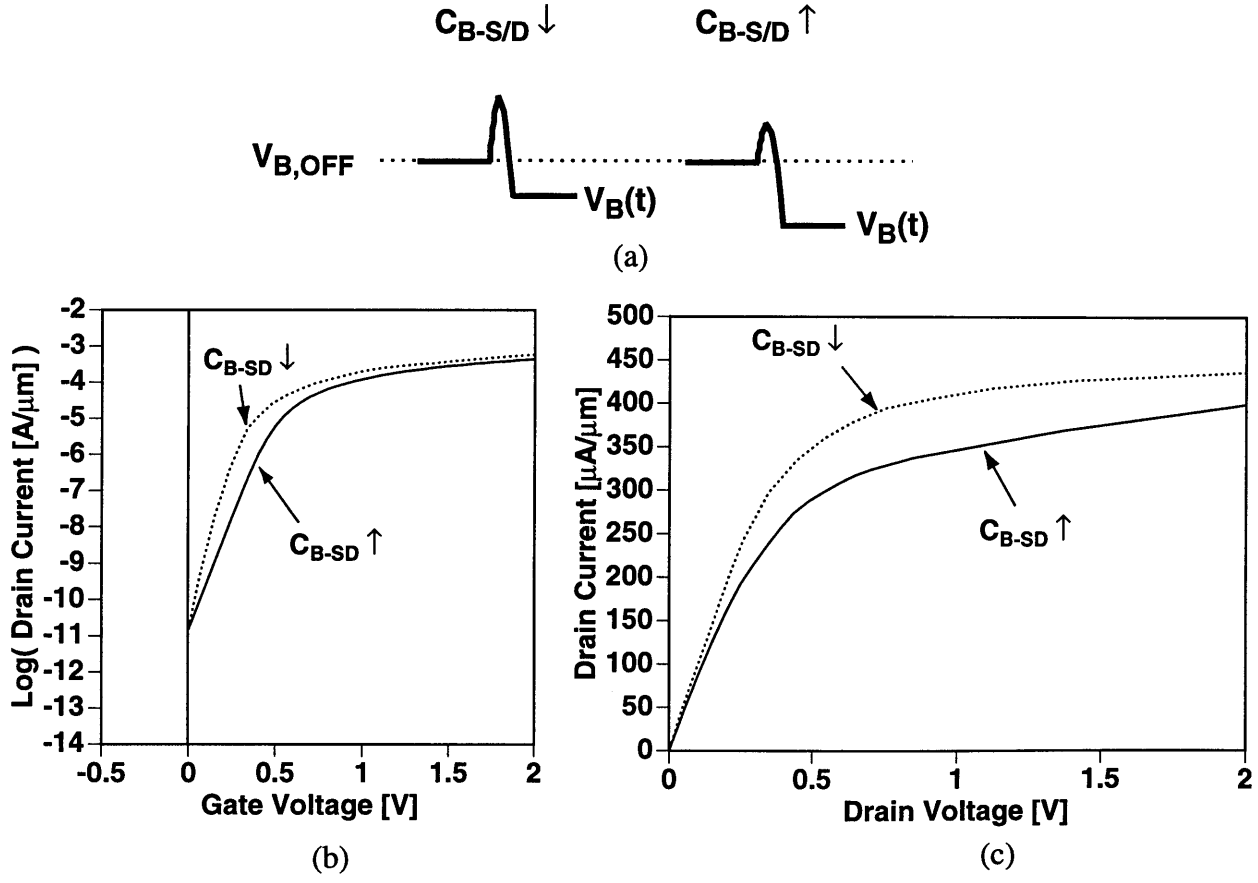


Figure 2-21. Schematic showing what happens to *I-V*'s when body-source/drain capacitance changes independent of $V_{B,OFF}$. Decreasing body-source/drain capacitance increases the gate-body current boost and decreases drain-body current drag. (a) shows the affect on the body voltage through a switch as defined in figure 2-16, (b) shows the effect on gate transfer *I-V*'s, and (c) shows the effect on the output characteristics.

when the drain falls. As shown in Figure 2-21 (b) and (c), this translates into steeper subthreshold slope (decreases S), lower V_T, higher I_{ON}, higher R_{OUT}, and lower R_{CH}.

Keeping $V_{B,OFF}$ low and reducing the body-source/drain capacitance tends to produce better MOSFET parameters. Thinning the silicon film and reducing the body and source/drain doping concentrations reduce the capacitances between the body-source/drain. In addition, reduction of body-source/drain doping concentrations tends to reduce $V_{B,OFF}$, further enhancing the subthreshold slope.

Unfortunately, MOSFET scaling trends run counter to many of the desirable design aspects listed above. As the gate length shrinks, the body-source/drain doping concentrations must increase to maintain electrostatic integrity. This tends to decrease the magnitude of gate-body capacitive coupling relative to body-source/drain capacitive coupling. The silicon film thickness is still a variable, but as the power supply decreases with MOSFET scaling, hysteretic floating-body effects become more significant and limit the range of silicon film thicknesses which can be used. This will be discussed in more detail in the next chapter.

2.4 Summary

This chapter showed the effect of the floating body on SOI MOSFET I-V's with respect to CMOS switching behavior. An equivalent model of the floating body was presented. This model was used to qualitatively analyze the effect of the floating body on NMOSFET switching behavior in a CMOS inverter (the PMOSFET behavior is complementary) in terms of MOSFET parameters I_{OFF}, S, V_T, I_{ON}, R_{OUT}, and R_{CH}.

These MOSFET parameters were shown to vary with device design due to the floating body. In summary, let us review how these MOSFET parameters vary as devices are designed in accordance with scaling trends. The shorter gate length requires higher body doping to maintain electrostatic integrity. Higher body doping increases diode currents and diode symmetry due to tunneling in body-source/drain diodes. The increase in diode symmetry tends to increase $V_{B,OFF}$

(however, reduction of V_{DD} will reduce $V_{B,OFF}$ slightly). Increased $V_{B,OFF}$ and reduced gate-body capacitance relative to body-source capacitance degrades the subthreshold swing S. Reduced body-gate capacitance relative to drain-body capacitance degrades R_{OUT}. Reduction of silicon film thickness can minimize the degradation of S and R_{OUT}, but hysteretic floating-body effects will limit the range of silicon film thicknesses which can be used.

Chapter 3

Hysteretic Floating-Body Effects

As discussed in the previous chapter, the equilibrium floating-body majority carrier content of a PD-SOI MOSFET depends on the applied gate, source, drain, and substrate terminal voltages. Rapid switching of terminal voltages induce nearly-negligible changes in body majority carrier content since body majority carriers are trapped by body-source/drain junction potential barriers. However, the change in body majority carrier content can become significant over many switching cycles, eventually reaching a "switching-steady-state" value if kept switching (different value for different switching patterns). However, it can just as well returning to the initial value if the terminal voltages are returned to the initial settings and enough time is allowed for to reach equilibrium. Or, if left at another set of terminal voltages for a long period of time, the body majority carrier content will reach another equilibrium value associated with that set of terminal voltages.

This "hysteretic" variation of the body majority carrier content is problematic because many of the MOSFET parameters discussed in the previous chapter change with changing body majority carrier content. In particular, hysteretic threshold voltage variation is of serious concern in SOI CMOS circuit design because it can lead to propagation delay variation, particularly as the power supply is reduced. This chapter focuses on hysteretic behavior of PD-SOI MOSFET's. Again, the

main example used will be the NMOS in a CMOS inverter. It will be shown that much of the threshold voltage variation can be minimized by proper design of the silicon film thickness and that device scaling trends tend to minimize hysteretic behavior.

3.1 Logic State Threshold Voltage Variation

The relationship between logic states, body hole content (N_H), preswitch body voltage ($V_{B,pre}$), and threshold voltage (V_T) is illustrated in Figure 3-1. Figure 3-1 (a) shows a cascade of two CMOS inverters in equilibrium with the input at 0 V. The first-stage inverter is in the OUT-HI logic state with its NMOS is in the OFF-state ready to switch ON. We replace the OFF-state body voltage, $V_{B,OFF}$, terminology used in chapter 2, with a more general term, the preswitch body voltage, $V_{B,pre}$, which is the floating-body voltage in the OFF-state just prior to switching ($V_{B,pre}=V_{B,OFF}$ for the first switch from equilibrium). In OFF-state equilibrium, the body hole content is determined by how many holes are depleted with $V_D=V_{DD}$, $V_G=V_S=0$ V, and $V_B=V_{B,pre}$, as shown in Figure 3-1 (b). Recall that $V_{B,pre}$ is the body voltage which forward biases the body-source junction enough to remove holes generated by impact ionization and in the reverse biased body-drain junction.

The second-stage inverter is in the OUT-LO logic state with its NMOS in ON-state equilibrium. The body hole content of this device is determined how many holes are depleted by $V_G=V_{DD}$, $V_D=V_S=V_B=0$ V as shown in Figure 3-1 (c). V_B is 0 V because the source and drain are both at 0 V. The equilibrium body hole content in the second-stage OUT-LO NMOS will most likely be different than that of the first-stage OUT-HI NMOS. When the input to the cascade of inverters rises, the first stage NMOS switches ON as the gate voltage rises. The drain voltage falls discharging the gate capacitance of the second stage, switching OFF the second stage NMOS.

The body hole content of the first- and second-stage NMOSFET's change a negligible amount during this rapid switch. The second-stage NMOS is now ready to switch ON. However, it has a different body hole content compared to the first-stage NMOS when it was ready to switch ON. Thus, the second-stage NMOS will switch ON differently than the first-stage NMOS as follows: if

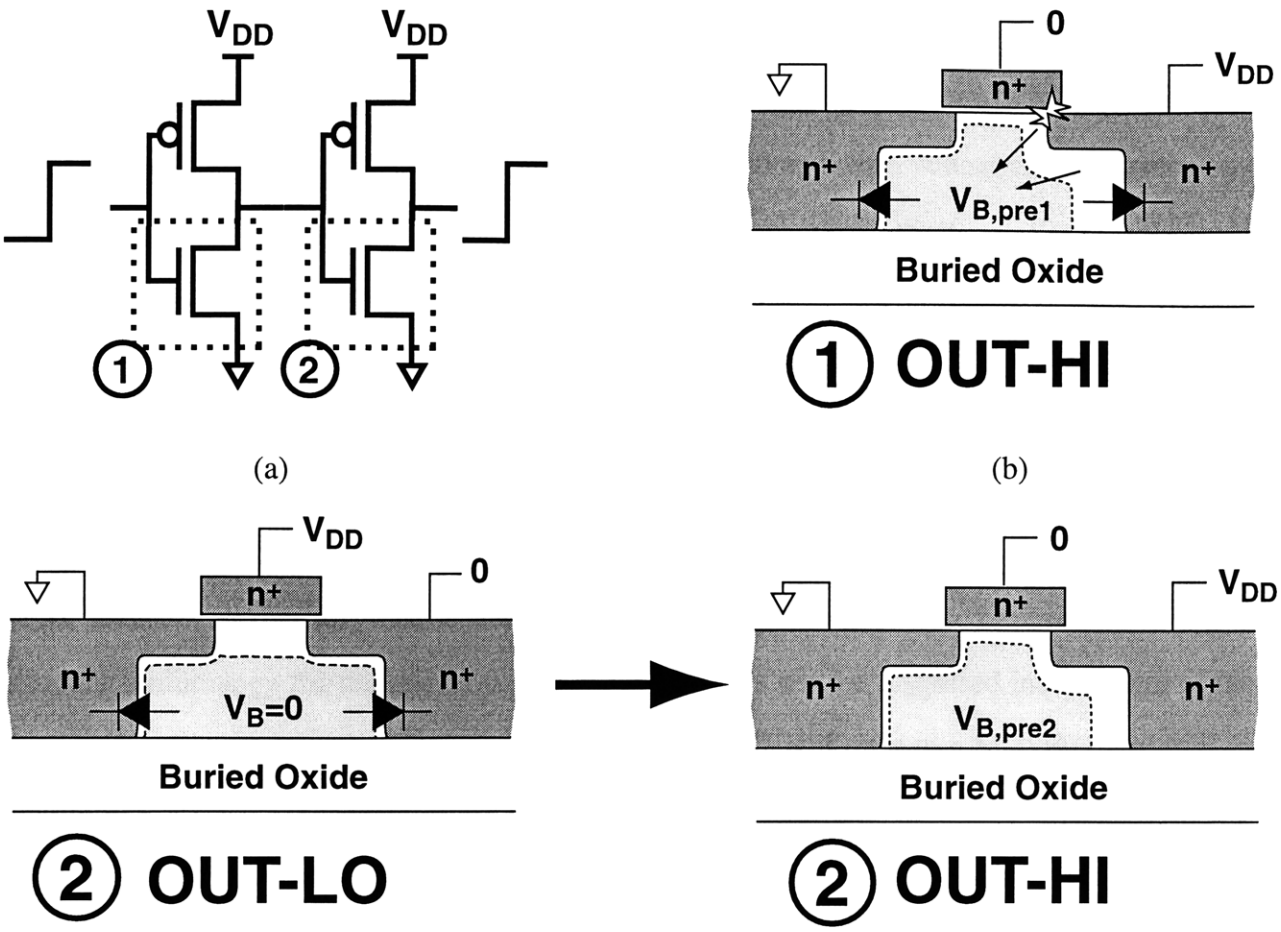


Figure 3-1. (a) Two stages in a CMOS invert chain, with NMOSFET's highlighted, in steady-state then switching from input-LO to input-HI. The first stage begins output-HI and switches to output-LO. The second stage begins in output-LO and switches to output-HI. (b) Schematic showing the preswitch voltage, $V_{B,pre1}$, and hole content in the first-stage NMOS in the OFF-state (OUT-HI), before switching-ON. (c) Schematic showing the second-stage NMOS hole content and body voltage in the ON-state (OUT-LO). When the first stage switches, the second-stage NMOS switches to the OFF-state (OUT-HI). Due to near-instantaneous switching, the second-stage NMOS maintains the body hole content it had in the ON-state (OUT-LO). The second-stage preswitch voltage, $V_{B,pre2}$, reflects the body hole content from the OUT-LO condition.

the body hole content in the equilibrium OUT-HI first-stage NMOS is greater/less than the body hole content in the equilibrium OUT-LO second-stage NMOS, then the first-stage NMOS will switch faster/slower than the second-stage NMOS due to it having a lower/higher threshold voltage.

An equally valid way to look at this is through the preswitch body voltage. The first-stage

OUT-HI NMOS has a preswitch body voltage, $V_{B,pre1}$, set by the diodes and impact ionization. The second-stage OUT-LO NMOS starts at V_B=0 V, but when it switches to the OUT-HI logic state, the body voltage is capacitively coupled to the gate voltage fall and drain voltage rise. The resulting body voltage is the preswitch body voltage for the second-stage NMOS, $V_{B,pre2}$. Because the body hole content between the first- and second-stage is different, $V_{B,pre1} \neq V_{B,pre2}$. Recall that the threshold voltage changes with $V_{B,pre}$. Thus, the second-stage NMOS will switch ON differently than the first-stage NMOS as follows: if $V_{B,pre1}$ is greater/less than $V_{B,pre2}$, then the first-stage NMOS will switch faster/slower than the second-stage NMOS due to it having a lower/higher threshold voltage.

Note that if the example under discussion began with input into the cascade of inverters HI rather than LO, all the comparative terms in the above analysis will need to be reversed. Thus, to simplify terminology for the rest of the thesis, MOSFET's will be discussed independent of the circuit with the following terminology: "starting OUT-HI" will refer to a MOSFET which is in equilibrium in a CMOS inverter in the OUT-HI logic state and "starting OUT-LO" will refer to a MOSFET which is in equilibrium in a CMOS inverter in the OUT-LO logic state. "OUT-HI" and "OUT-LO" will be used to denote the current logic state of the CMOS inverter which contains the MOSFET(s) under discussion. With this terminology we can say that for a given NMOSFET, the threshold voltage starting OUT-HI is lower/higher than starting OUT-LO if the body hole content starting OUT-HI is greater/less than starting OUT-LO. Thus, the NMOSFET starting OUT-HI will switch faster/slower than starting OUT-LO. However, a MOSFET starting OUT-LO must switch OUT-HI before it can be compared to a MOSFET starting OUT-HI, as is shown in Figure 3-1 (b)-(d).

3.2 Silicon Film Thickness

Figure 3-2 shows why a MOSFET starting OUT-HI most likely has different body hole content compared to starting OUT-LO. It is simply because deletion regions starting OUT-HI are different than the depletion regions starting OUT-LO, usually resulting in a different body hole

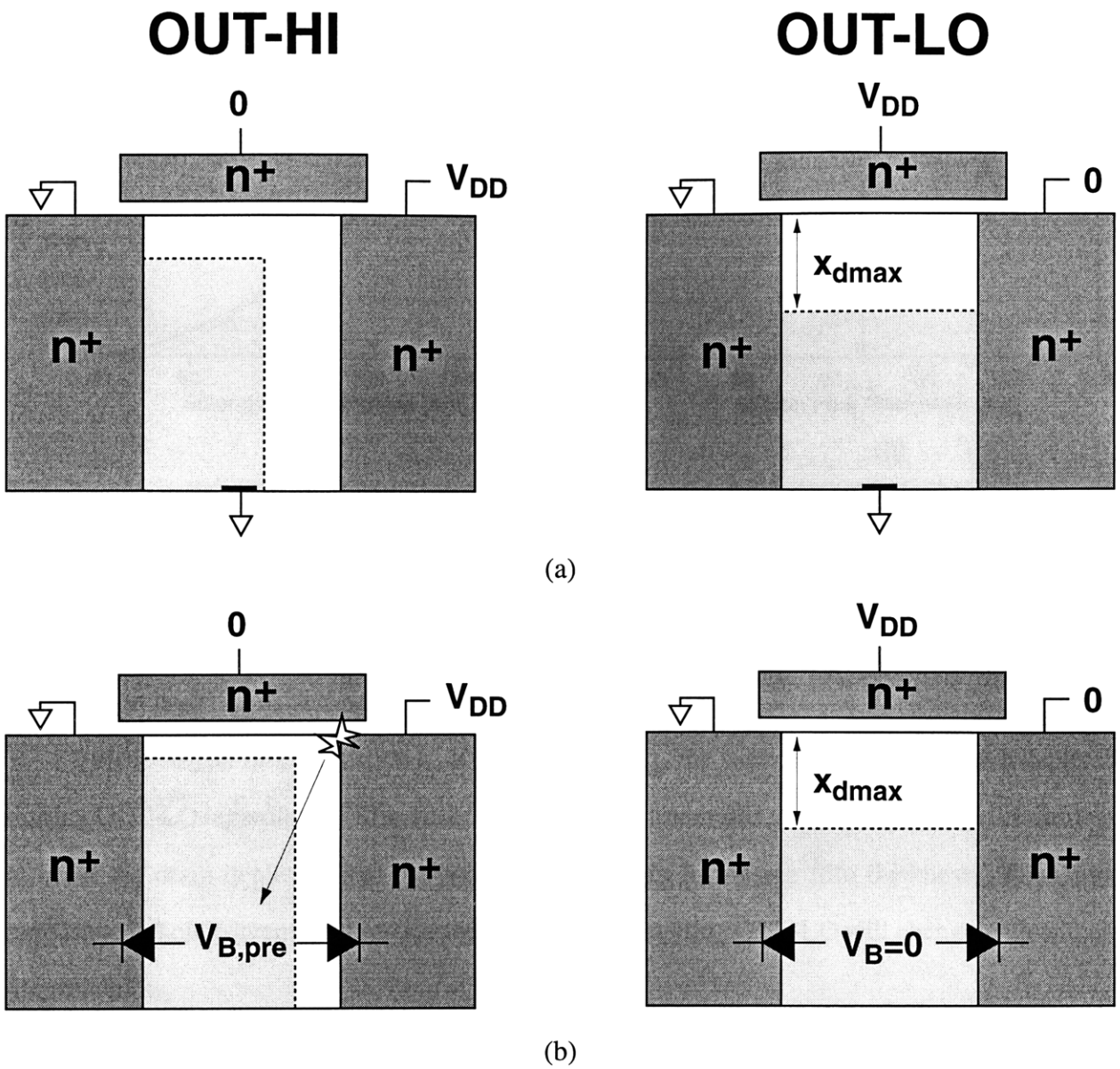


Figure 3-2. Schematic showing major NMOS body depletion regions in OUT-LO and OUT-HI logic states. (a) Body-contacted devices are shown for reference. The floating-body case is shown in (b). In OUT-HI, there are more holes than when body-contacted due to $V_B=V_{B,pre}>0$ shrinking the depletion regions. OUT-LO is the same as the body-contacted case since the body voltage floats to 0 V.

content as shown for a NMOSFET in Figure 3-2. Starting OUT-HI, body hole depletion is due to the workfunction difference between the n$^+$-doped gate polysilicon and p-doped body and junction depletion of the body from the drain biased at V_{DD}. Starting OUT-LO, body hole depletion is due to the gate voltage at V_{DD}, which depletes the body to a maximum depletion depth x_{dmax}, at which point

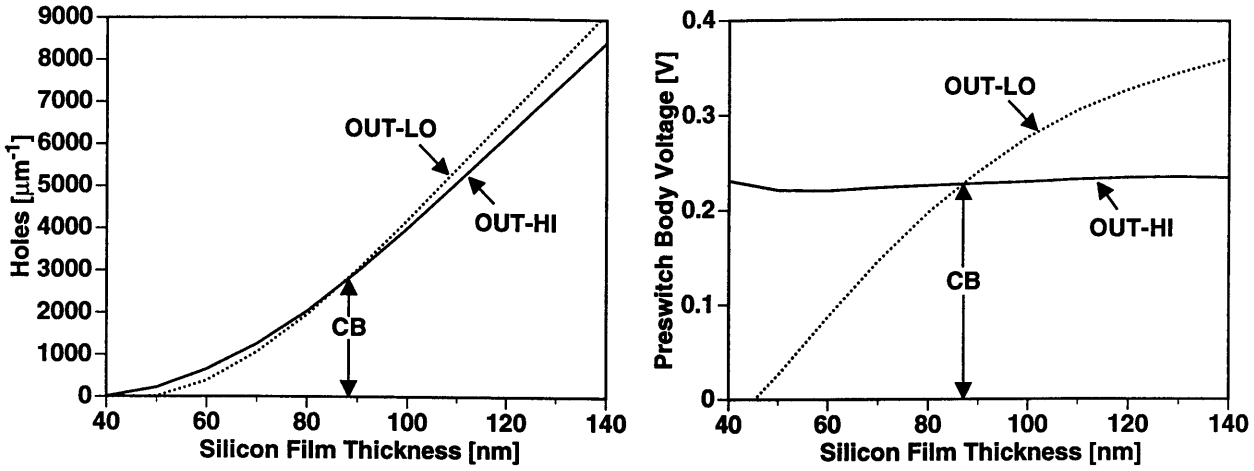
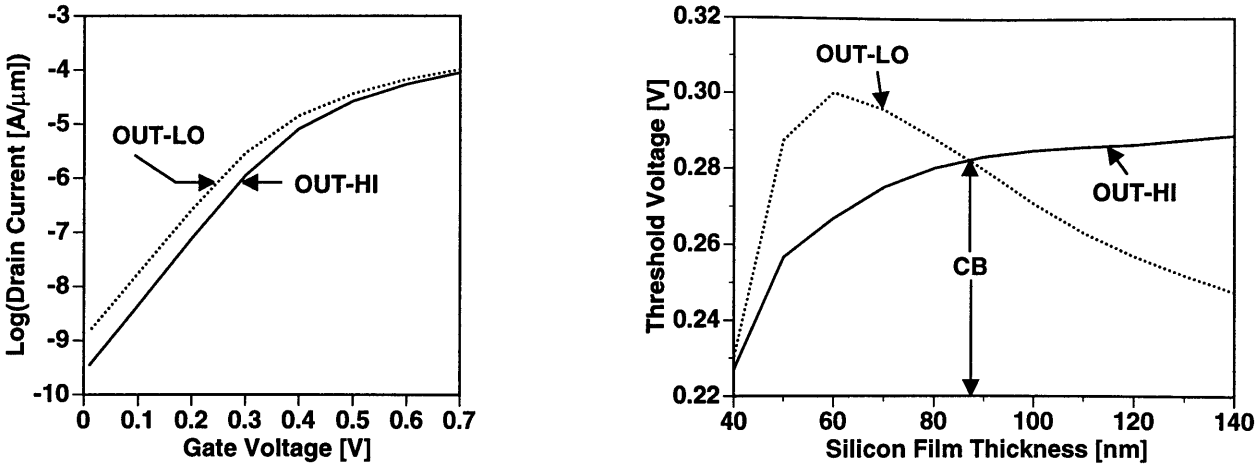


Figure 3-3. (a) Body hole content and (b) preswitch body voltage versus silicon film thickness for a NMOSFET starting OUT-HI and starting OUT-LO. Note that the silicon film thickness where OUT-HI and OUT-LO are equal denotes "charge-balance".

an inversion layer forms at the surface.

If the silicon film thickness, t_{si}, is greater than x_{dmax}, the same number of holes are depleted starting OUT-LO regardless of film thickness. This is not true starting OUT-HI, because the number of holes the drain depletes from the body increases with increasing film thickness. Thus, any imbalance in hole content between starting OUT-HI and starting OUT-LO will change with silicon film thickness.

This is shown in more detail in Figure 3-3 (a) which compares the body hole content starting OUT-HI and starting OUT-LO for NMOSFET with L_{GATE}=0.18 μm, L_{eff}=0.12 μm, t_{ox}=4.5 nm, body doping N_A=4x10^{17} cm^{-3}, t_{box}=200 nm, and V_{DD}=1.5 V. Note that x_{dmax} is 50 nm in this example. At thinner films, the body hole content is greater starting OUT-HI than starting OUT-LO. The body hole content difference changes with increasing silicon film thickness, converging to a point at which the body hole content is equal starting OUT-HI and starting OUT-LO. This point is called the "charge-balance" silicon film thickness. At silicon film thicknesses above the charge-balance thickness, $t_{si,CB}$, the body hole content starting OUT-HI becomes less than starting OUT-LO. Figure 3-3 (b) shows that the preswitch body voltage exhibits the same behavior, as should be expected



(a) (b)

Figure 3-4. (a) Subthreshold gate transfer characteristics for a NMOSFET starting OUT-HI and starting OUT-LO. In this example t_{si}=140 nm. (b) Threshold voltage versus silicon film thickness for a NMOSFET starting OUT-HI and starting OUT-LO. Note that the silicon film thickness where OUT-HI and OUT-LO are equal denotes "charge-balance".

since the body hole content is directly related to the preswitch body voltage by the AC capacitive network in Figure 2-7 (b).

When there is a difference in the body hole content starting OUT-HI versus starting OUT-LO, there will be a difference in threshold voltage. This is shown in Figure 3-4. 3-4 (a) shows an example of the gate transfer *I-V* characteristics for the NMOSFET structure under discussion with t_{si}=140 nm. Recall from Figure 3-3 that the body hole content and preswitch body voltage starting OUT-LO is greater than starting OUT-HI at t_{si}=140 nm. This translates to the threshold voltage starting OUT-LO being less than starting OUT-HI, as shown in Figure 3-4 (a). The threshold voltage versus film thickness for this structure starting OUT-HI and starting OUT-LO is shown in Figure 3-4 (b). Indeed, the threshold voltage starting OUT-HI is equal to starting OUT-LO at the charge-balance silicon film thickness. Note that the bigger the difference in body hole content between starting OUT-HI and starting OUT-LO, the larger difference in threshold voltage. Thus, design at or near the charge-balance silicon film thickness minimizes threshold voltage variation with logic-state. (Also note that between t_{si}=50 nm to 40 nm, the threshold voltage begins to

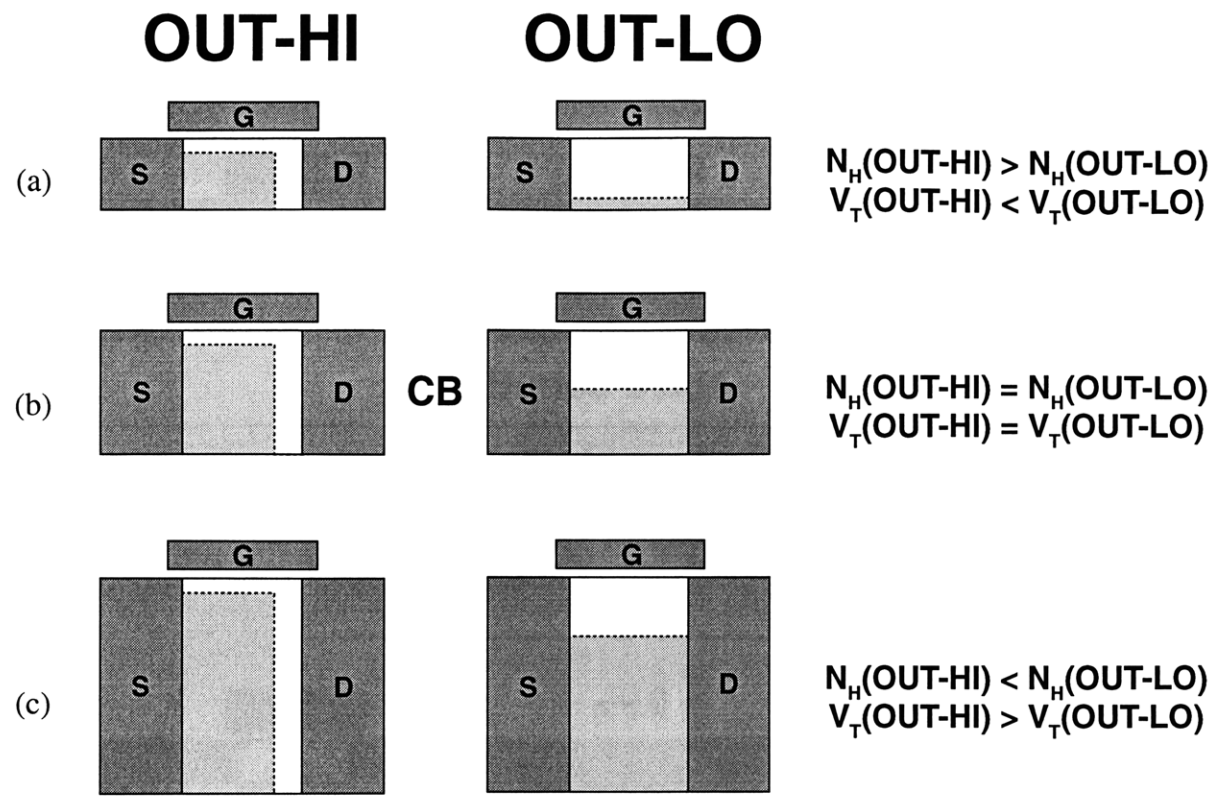


Figure 3-5. Schematic showing NMOS body hole content trend versus silicon film thickness. For a given gate length and power supply voltage, the silicon film thickness determines whether (a) OUT-HI has more holes than OUT-LO, (b) OUT-HI and OUT-LO are charge-balanced, or (c) OUT-HI has fewer holes than OUT-LO.

converge because MOSFET is in transition between partially-depleted and fully-depleted).

Figure 3-5 summarizes the body hole content and threshold voltage trend versus silicon film thickness. For silicon film thickness less than the charge-balance thickness, the body hole content starting OUT-HI is greater than starting OUT-LO, the threshold voltage starting OUT-HI is less than starting OUT-LO. For silicon film thickness greater than the charge-balance thickness, the body hole content starting OUT-HI is less than starting OUT-LO, the threshold voltage starting OUT-HI is greater than starting OUT-LO. Recall that lower threshold voltage will result in a "faster" switching device, whereas a higher threshold voltage will result in a "slower" switching device.

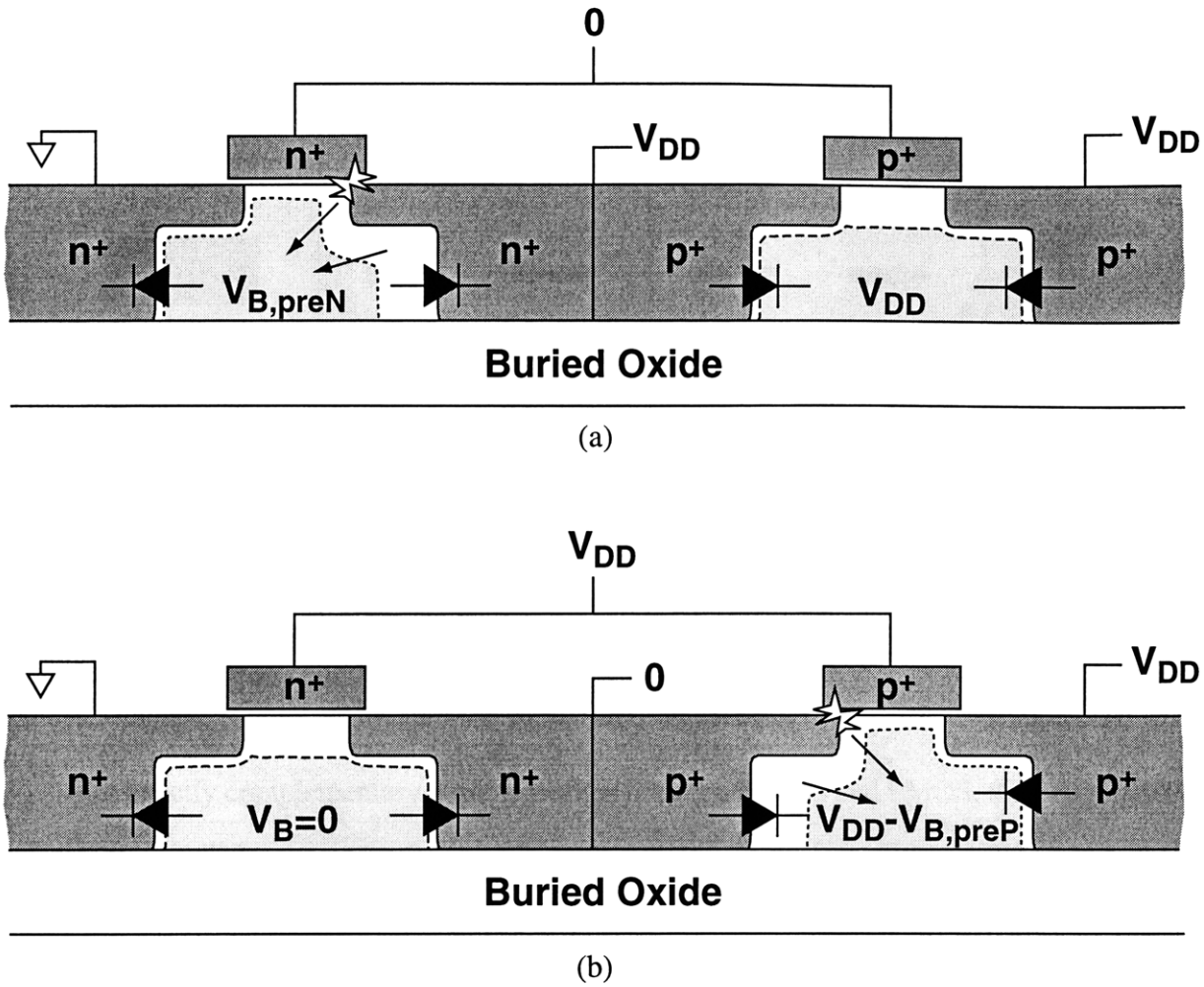


Figure 3-6. Schematic showing NMOS and PMOS equilibrium body majority carrier content in a CMOS inverter with (a) OUT-HI and (b) OUT-LO. NMOS and PMOS are complementary.

3.3 NMOS versus PMOS

As usual, NMOS and PMOS behave in a complementary manner in a CMOS inverter. As shown in Figure 3-6, the NMOS body depletion regions starting OUT-HI are similar to the PMOS body depletion regions starting OUT-LO because these devices are in the OFF-state. The NMOS body depletion regions starting OUT-LO are similar to the PMOS body depletion regions starting OUT-HI, because these devices are in the ON-state.

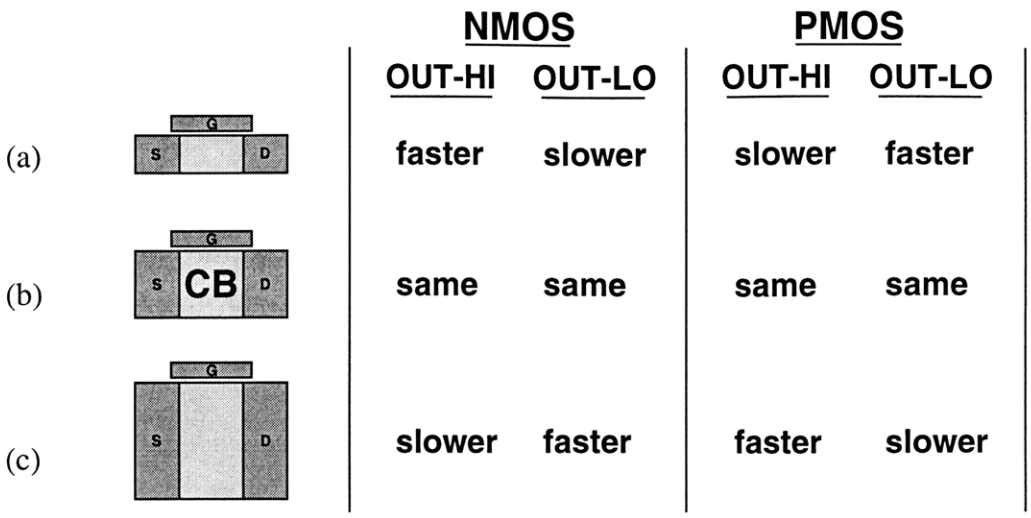


Figure 3-7. Relative device switching speed starting from equilibrium with silicon film thickness (a) less than charge-balanced, (b) charge-balanced, and (c) thicker than charge-balanced for NMOS and PMOS. Note that NMOS charge-balanced silicon film thickness may be different than PMOS charge-balanced silicon film thickness.

For exactly complementary doping profiles between NMOS and PMOS, the body majority carrier content will be the same for the ON-state devices (NMOS starting OUT-LO and PMOS starting OUT-HI). However, the body majority carrier content may be different in the OFF-state devices (less in PMOS starting OUT-LO compared to NMOS starting OUT-HI) due to much reduced OFF-state impact ionization generation in PMOSFET's. This results in the PMOS having less body majority carrier content in the OFF-state compared to NMOS (equivalently, $V_{B,preP}$ shown in Figure 3-6 will be less than $V_{B,preN}$). Thus, the silicon film thickness for charge balance will tend to be less than the silicon film thickness for charge balance for NMOS.

As devices are scaled, OFF-state impact ionization will be mitigated by increased reverse diode tunneling current due to the high peak body dopings required to maintain electrostatic integrity. Also, the power supply may the reduce the peak electric field resulting in lower impact ionization generation. Thus, device scaling trends will make simultaneous charge balance of NMOSFET's and PMOSFET's more probable. Figure 3-7 summarizes relative NMOS speed for starting OUT-HI and starting OUT-LO versus silicon film thickness relative to the NMOS charge balance film thickness, and relative PMOS speed for starting OUT-HI and starting OUT-LO versus silicon film

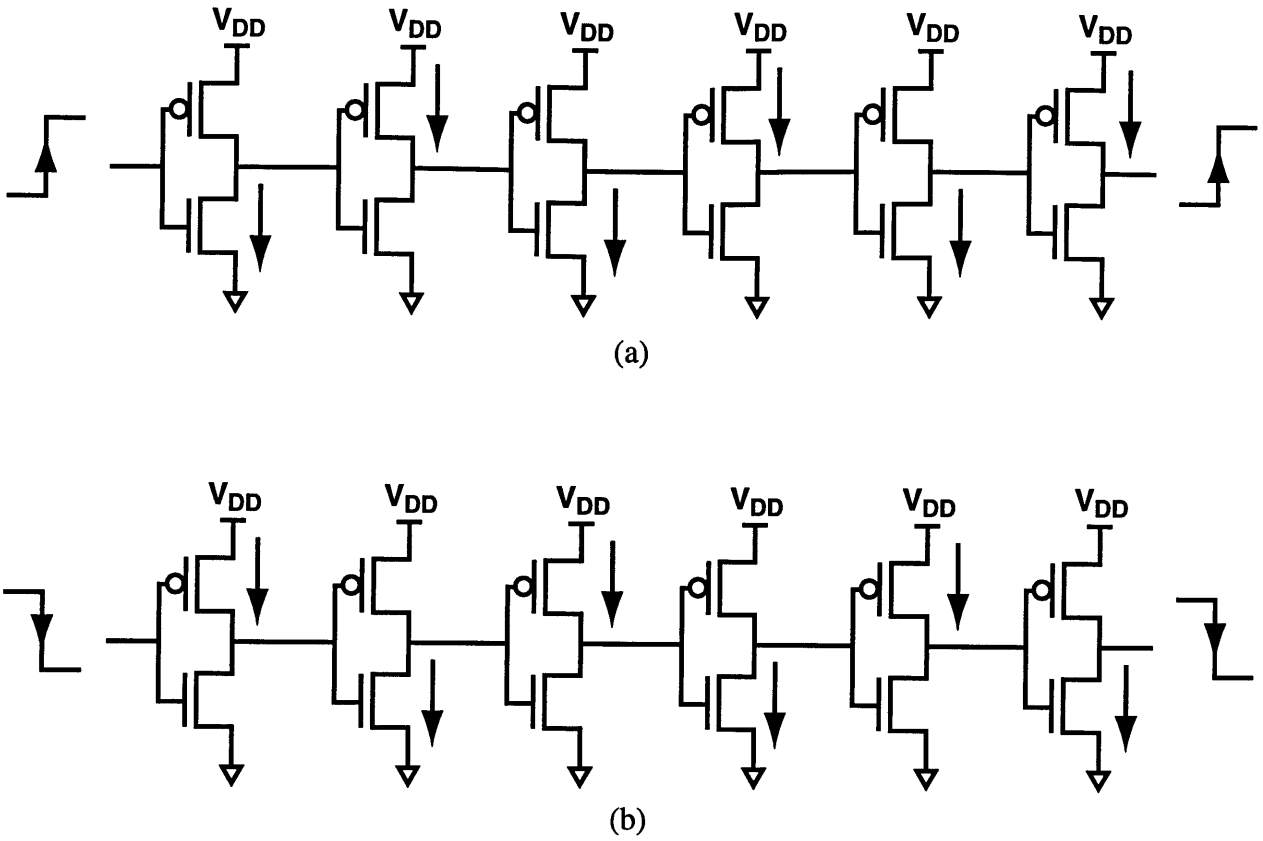


Figure 3-8. CMOS inverter chain showing which devices propagate the (a) rising edge of the input pulse and (b) the falling edge of the input pulse.

thickness relative to the PMOS charge balance film thickness.

3.4 Pulse Stretching

Figure 3-7 showed that due to the floating body, devices may be faster or slower depending on starting logic-state. This will affect propagation delays through circuits because different parts of an input signal propagate through groups of devices which may be collectively faster or slower. Figure 3-8 shows that the rising edge of a pulse into a CMOS inverter propagates through the NMOSFET's which start OUT-HI and PMOSFET's which start OUT-LO. The falling edge propagates through NMOSFET's which start OUT-LO and PMOSFET's which start OUT-HI. Thus, starting from equilibrium with the input LO and with MOSFET's where the silicon film thickness

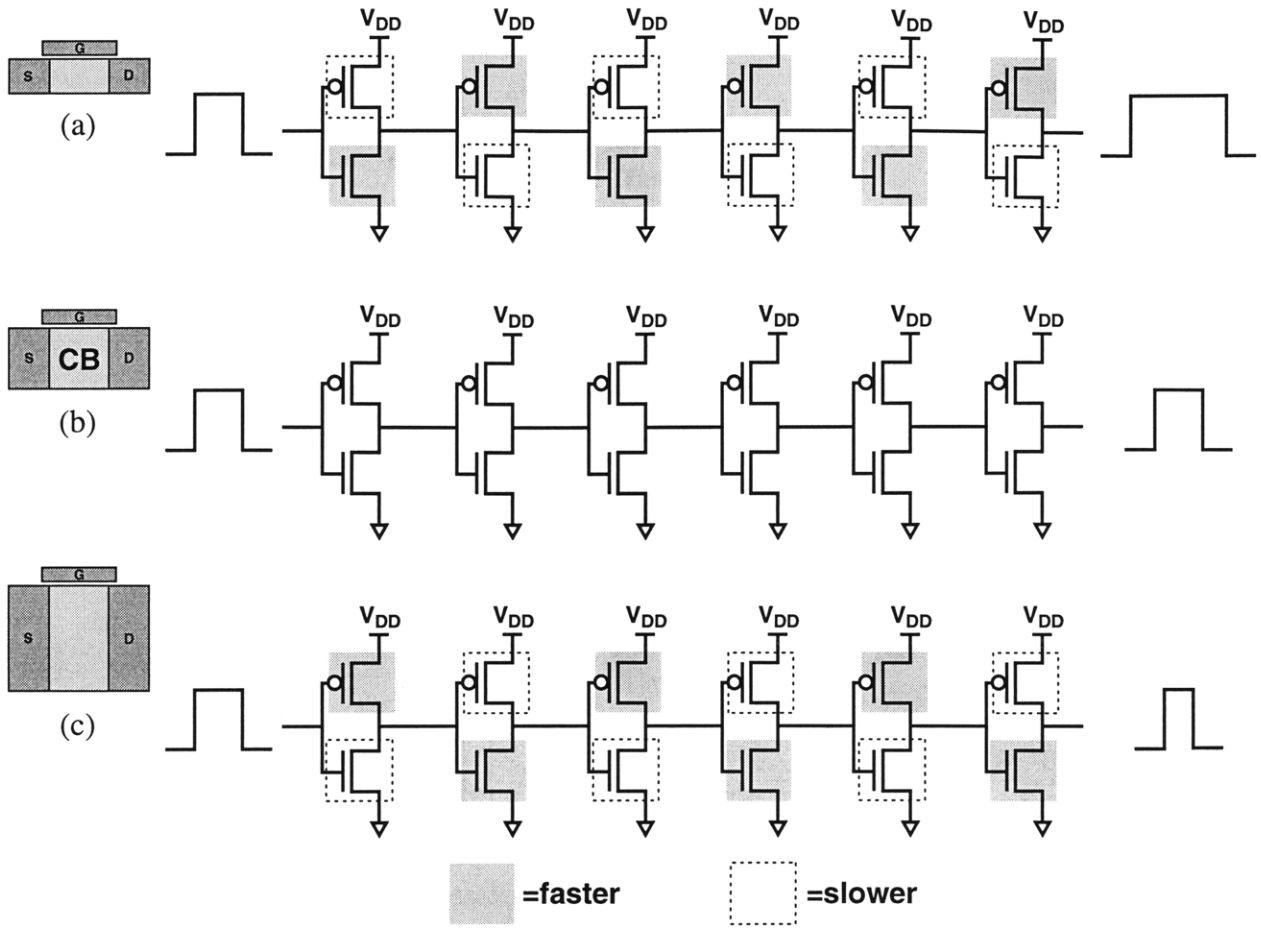


Figure 3-9. Output pulse stretching versus silicon film thickness relative to charge-balance. This figure assumes that all devices are in equilibrium prior to the input pulse. (a) The output pulse stretches if both NMOS and PMOS silicon film thicknesses are less than the charge-balanced thickness. (b) The output pulsewidth is equal to the input pulsewidth if NMOS and PMOS are both charge-balanced. (c) The output pulse compresses if both NMOS and PMOS silicon film thicknesses are greater than charge-balanced. Given enough stages, the pulse will cease to propagate and disappear. Note that NMOS charge-balanced silicon film thickness may be different than PMOS charge-balanced silicon film thickness.

is less than charge-balanced, the rising edge of the input pulse propagate through devices which are collectively faster. The falling edge of the input pulse propagates through devices which are collectively slower. The result is that the rising edge will arrive at the output faster than expected and the falling edge will arrive at the output slower than expected - the output pulse will "stretch".

If both NMOS and PMOS are charge-balanced, then the output pulse will have the same width as the input pulse due to lack of logic-state dependent speed variation. If the silicon film

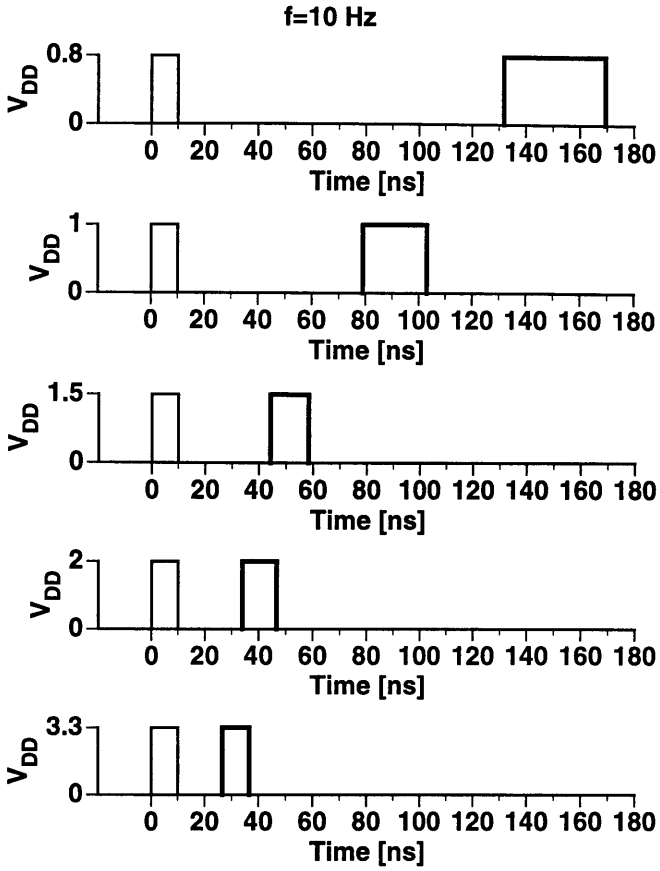


Figure 3-10. Pulse stretching measured in a 480-stage CMOS inverter chain at various supply voltages. The input pulse is 10 ns wide. The output pulse is shown in bold. The input pulse frequency is 10 Hz. This low frequency is equivalent to starting from equilibrium since there sufficient time between pulses to restore equilibrium.

thickness of the devices in the CMOS inverter chain are greater than the charge-balance thickness, then the rising edge will arrive at the output slower than expected and the falling edge will arrive at the output faster than expected - the output pulse will "compress". If the chain is long enough, the pulse will disappear because at some point the rising edge slows down enough and the falling edge speeds up enough for the transition between rising and falling edge to be too rapid for the CMOS inverters to respond. Figure 3-9 summarizes pulse stretching trends versus silicon film thickness relative to both NMOS and PMOS being charge-balanced.

The pulse stretching effect was measured at Digital Equipment Corporation in a 480-stage inverter chain with L_{GATE}=0.5 μm, L_{eff}=0.35 μm, t_{ox}=8.5 nm, t_{si}=125 nm, N=4x10^{17} cm^{-3}, and t_{box}=380 nm [36]. The aspect ratio of the gate length to silicon film thickness suggests that the silicon film

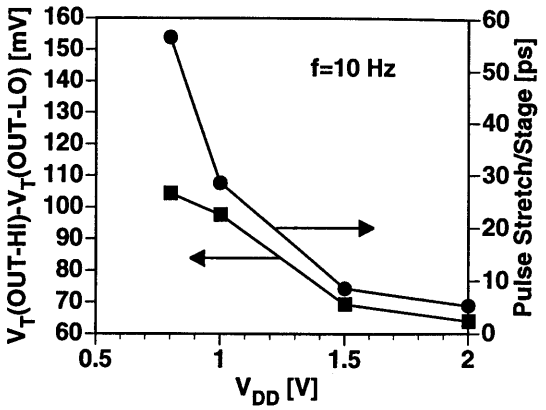


Figure 3-11. Measured pulse stretching per stage for a 480-stage PD-SOI CMOS inverter chain versus supply voltage at input pulse frequency of 10 Hz. Also plotted is the initial difference in threshold voltage between slower and faster devices, as extracted from SPICE simulation assuming nominal $|V_T|$=0.3 V. Note that devices are more charge-balanced at higher supply voltage.

thickness is less than the charge-balance thickness. Therefore, as Figure 3-10 shows, the output pulse stretches relative to the input pulse, particularly at low supply voltages. As V_{DD} increases, pulse stretching decreases for two reasons. First, higher V_{DD} is less sensitive to threshold voltage variation, ΔV_T, since $I_{DSAT} \alpha (V_{DD}-V_T)$. Second, as V_{DD} increases, the devices become more charge-balanced because the body majority carriers depleted by the drain increases with V_{DD}, whereas the body majority carriers depleted by the gate is fixed at x_{dmax} times the doping, regardless of V_{DD}. This decrease in ΔV_T with increasing V_{DD} is shown in Figure 3-11 which extracts ΔV_T between starting OUT-HI and starting OUT-LO from SPICE simulation of the inverter chain under discussion.

3.5 Frequency Dependence

The inverter chain reaches a switching-steady-state for each switching pattern. The input pulse frequency shown in Figure 3-10 was 10 Hz. This low frequency is equivalent to starting from equilibrium since there is sufficient time between pulses to restore equilibrium. Thus, the switching-steady-state for this switching waveform is equivalent to equilibrium. As the input pulse frequency increases however, there will not be sufficient time between pulses to restore equilibrium. Therefore, propagation delays change with input pulse frequency, changing the amount of pulse stretch. This

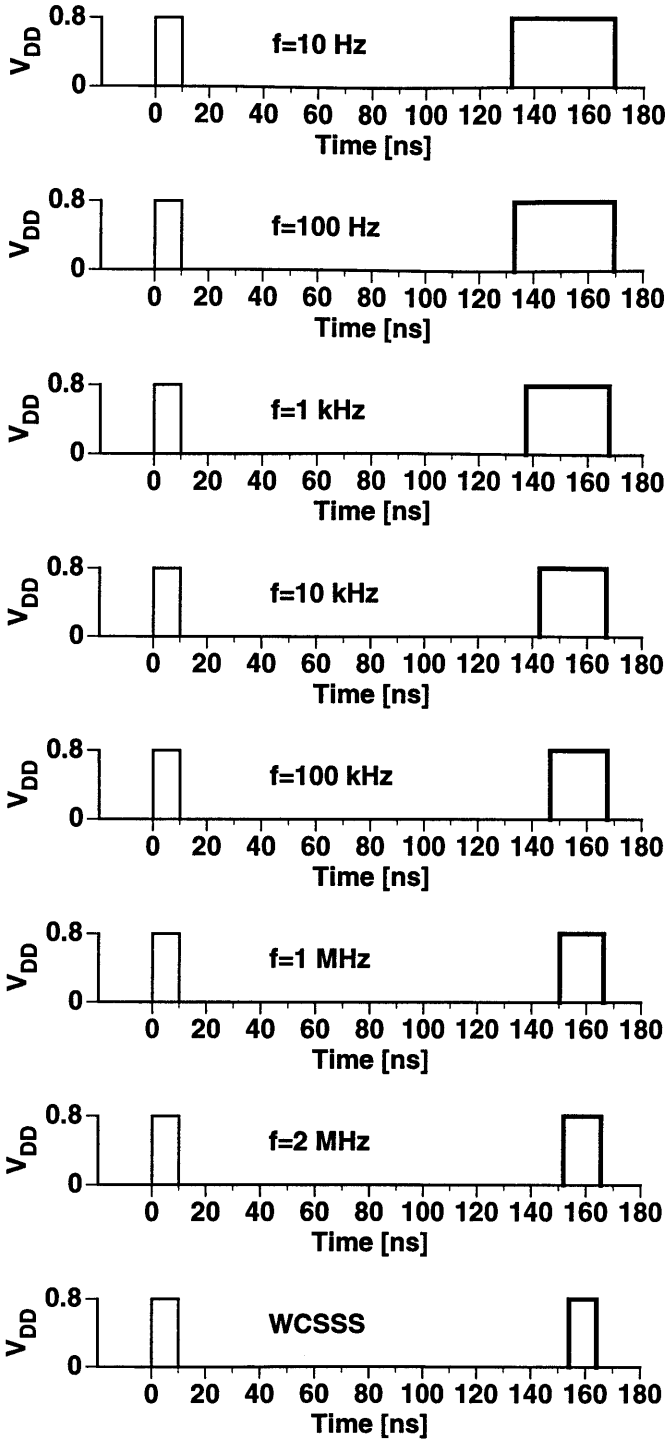


Figure 3-12. Pulse stretching measured in a 480-stage CMOS inverter chain for V_{DD}=0.8 V at various input pulse frequencies. The input pulse is 10 ns wide. The ouput pulse is shown in bold. The output pulse stretch decreases at higher frequencies. Measurements only up to 2 MHz were made and switching-steady-state has not been reached. A possible worst-case switching-steady-state is shown as reference.

is shown for the measured inverter chain at V_{DD}=0.8 V at various frequencies. As the input pulse frequency increases, the rising edge of the output pulse slows down and the falling edge speeds up. This indicates that the faster devices in the inverter chain slow down by losing body majority carriers and the slower devices in the inverter chain speed up by gaining body majority carriers.

The asymmetry in the slow down of the rising edge to the speed up in the falling edge indicates that it is easier to lose body majority carriers than it is to gain body majority carriers. This should be no surprise because the forward bias diode current responsible for loss of body majority carrier content is typically much greater than the reverse bias diode current responsible for gain in body majority carrier content. Impact ionization current is also responsible for gain in body majority carrier content, but the example shown in Figure 3-12 was measured at V_{DD}=0.8 V where impact ionization should be minimal for both NMOS and PMOS. At higher supply voltages, it is possible for both initially slower and faster NMOSFET's to speed up if impact ionization generation is significant.

As shown schematically in Figure 3-12, at high enough frequency the output pulse will no longer be wider than the input pulse frequency. This condition is defined as "worst-case switching-steady-state" and is a result of all the NMOSFET's having the same body hole content, regardless of starting OUT-HI or OUT-LO, and all PMOSFET's having the same body electron contents, regardless of starting OUT-HI or OUT-LO. Recall from Figure 3-9 that this results in zero pulse stretching. Therefore, the maximum range of threshold voltage variation is between the threshold voltage either starting OUT-HI or OUT-LO, to the threshold voltage at worst-case switching-steady-state.

Figure 3-13 shows that the frequency at which worst-case switching-steady-state is established decreases with increasing supply voltage. This is shown in more detail in Figure 3-14, which shows the rising and falling edges of the output pulse relative to the midpoint in the output pulse. As the supply voltage is increased, there is more forward diode current per switch period to decrease the body majority carrier content and more reverse diode current per switch period to increase body majority carrier content. This is due to capacitive coupling of the body voltage to the

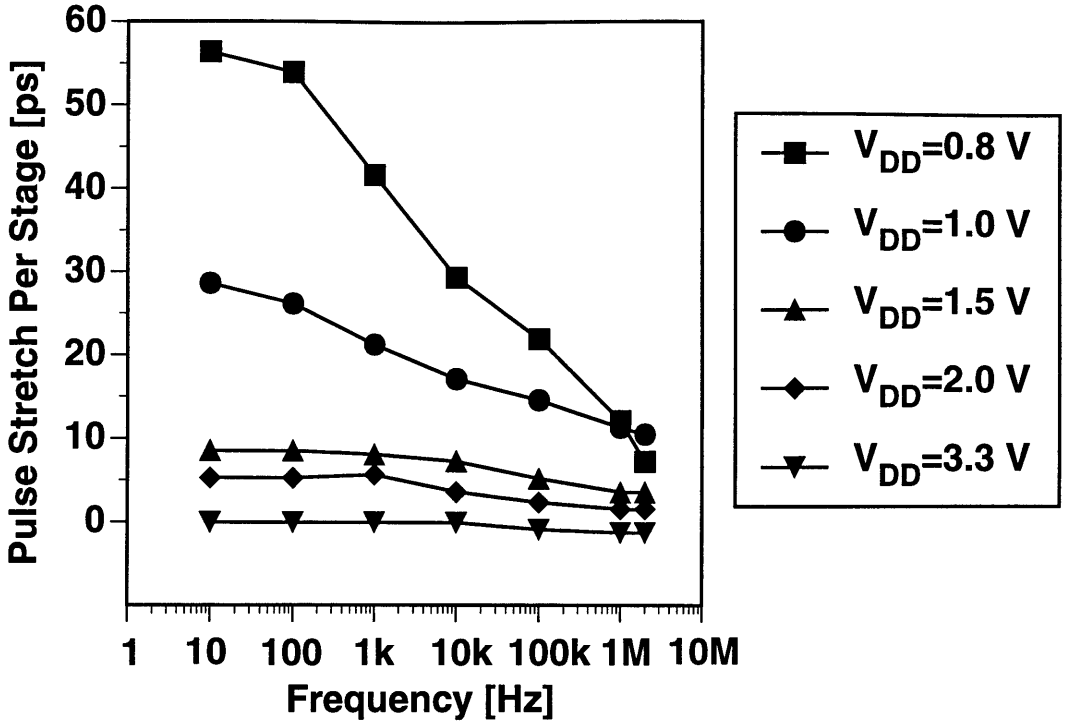


Figure 3-13. Pulse stretching per stage measured in a 480-stage CMOS inverter chain versus input pulse frequency at various supply voltages. The input pulse is 10 ns wide. Note that switching-steady-state (zero pulse stretch when output pulse width equals input pulse width of 10 ns) occurs at lower frequencies as supply voltage increases.

terminal voltages – the higher the terminal voltages, the higher the changes in body voltage, and hence the higher levels of body-source/drain diode currents. As noted above, impact ionization currents, particularly in NMOSFET's, increase exponentially with higher supply voltage, and higher levels help establish worst-case switching-steady-state at lower input pulse frequencies.

3.6 Hysteresis versus Device Design

Figures 3-15 and 3-16 look in more detail at what occurs in the NMOSFET of a CMOS inverter during switching. The figures show results of MEDICI 2-D numerical simulation for two sample devices switching as if in a CMOS inverter. The only difference between the two devices is the silicon film thickness. However, it will be shown that such slight changes in device design can have a great affect on the response of the floating body to switching. Figure 3-15 shows a NMOSFET

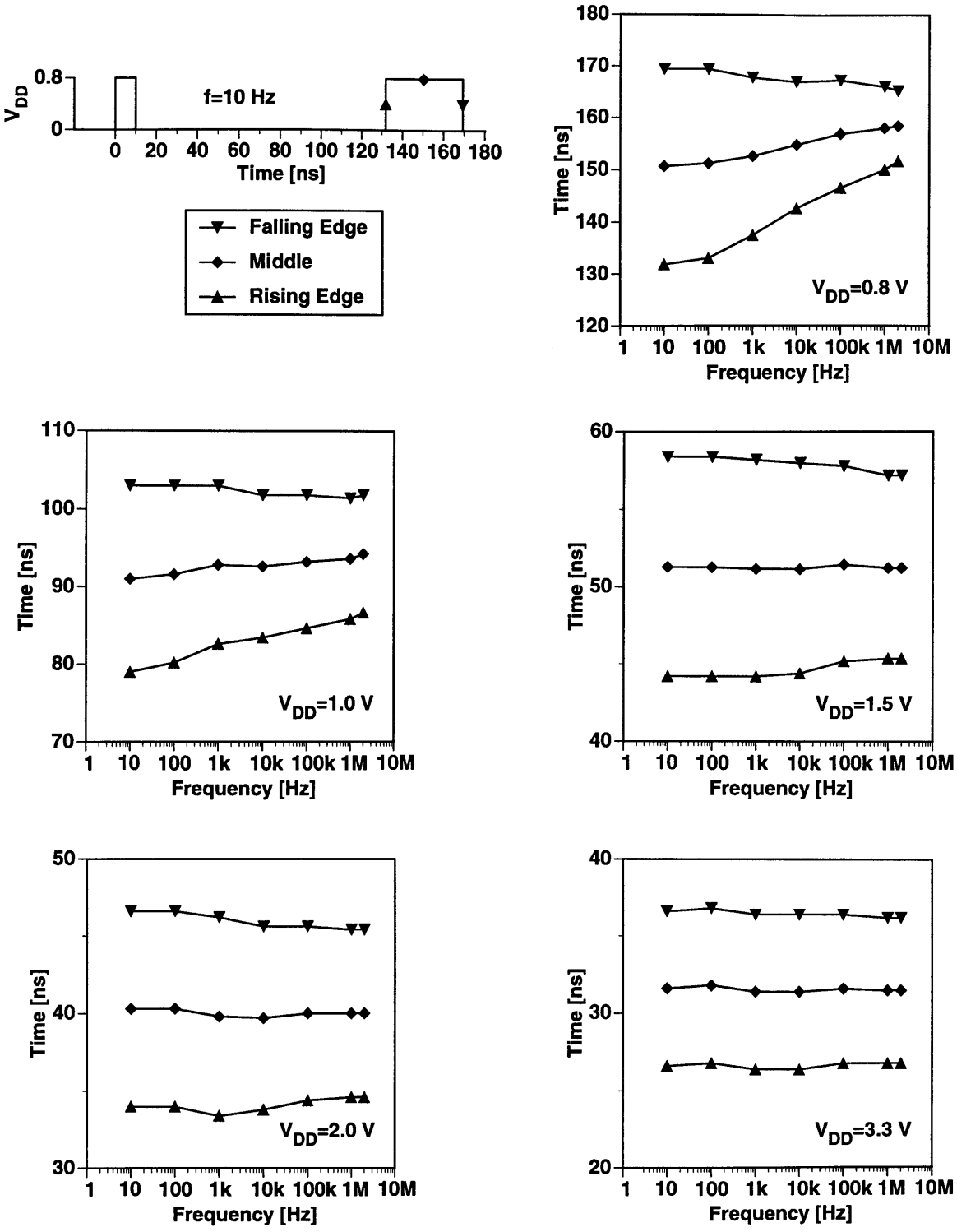
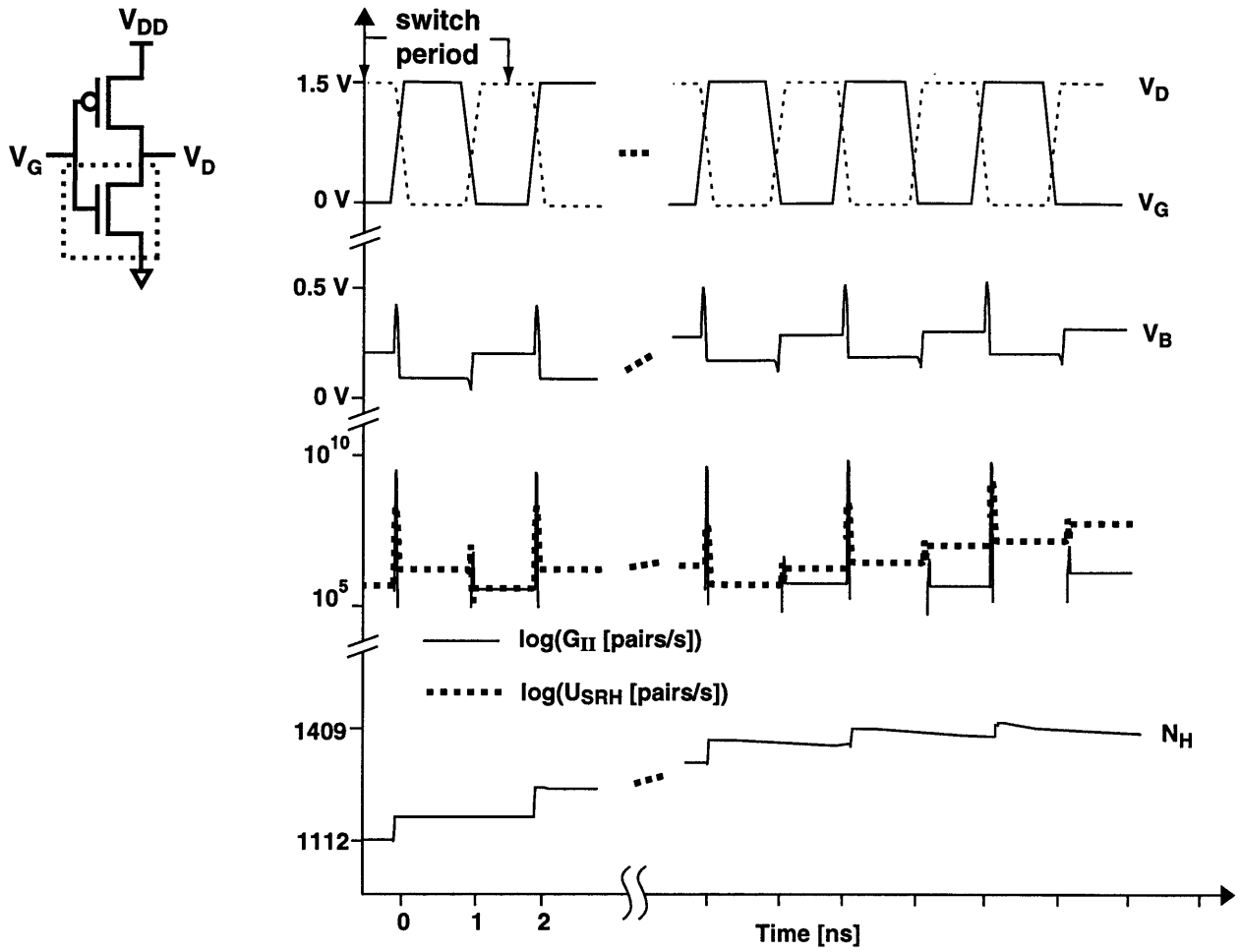


Figure 3-14. Output pulse rising edge, midpoint, and falling edge, measured in a 480-stage CMOS inverter chain versus input frequency at various supply voltages. The input pulse is 10 ns wide. Note that the rising edge slows down more than the trailing edge speeds up.

with t_{si}=70 nm, which is thinner than charge-balanced, and Figure 3-16 shows a NMOSFET with t_{si}=210 nm, which is thicker than charge-balanced.

Shown in Figure 3-15 (a) and (b) are the gate and drain waveforms as if the CMOS inverter is switching between OUT-HI and OUT-LO at 50 % duty cycle over a period of 2 ns. The fall and rise times of the gate and drain voltage waveforms is 50 ps, with a 40 ps delay between the gate and drain transitions. Also shown is the floating body voltage, the net recombination rate throughout the device on a logrithmic scale, U_{SRH} (U_{SRH}=R-G, where R is the thermal recombination rate throughout the structure, including recombination at the source/drain contacts, and G is the thermal generation rate), the impact ionization generation rate on a logrithmic scale, G_{II}, and the body hole content, N_H. Starting OUT-HI, the body voltage $V_{B,pre}$ is greater than zero. As the gate voltage rises, the body voltage is pulled up by capacitive coupling, increasing the forward bias on the body-source diode, which removes holes from the body. At the same time, drain current flows causing impact ionization generation of holes. Impact ionization generation is greater than body-source diode recombination during this transition so there is an increase in the body hole content. Now the NMOSFET is in OUT-LO for half a period. Note that the body voltage is pulled down, but not below 0 V. Thus the body-source and body-drain diodes are now forward biased, although with little recombination of the excess holes generated during the transition from OUT-HI to OUT-LO. The NMOSFET then switches back to OUT-HI. $V_{B,pre}$ has increased due to the increase in body hole content. As switching continues, the body hole content continues to increase during each switching period until the body voltage is sufficiently high that forward bias of the body-source/drain diodes over one switching period is sufficient to remove the holes generated by impact ionization. It may take many thousands of switches to reach this switching-steady-state.

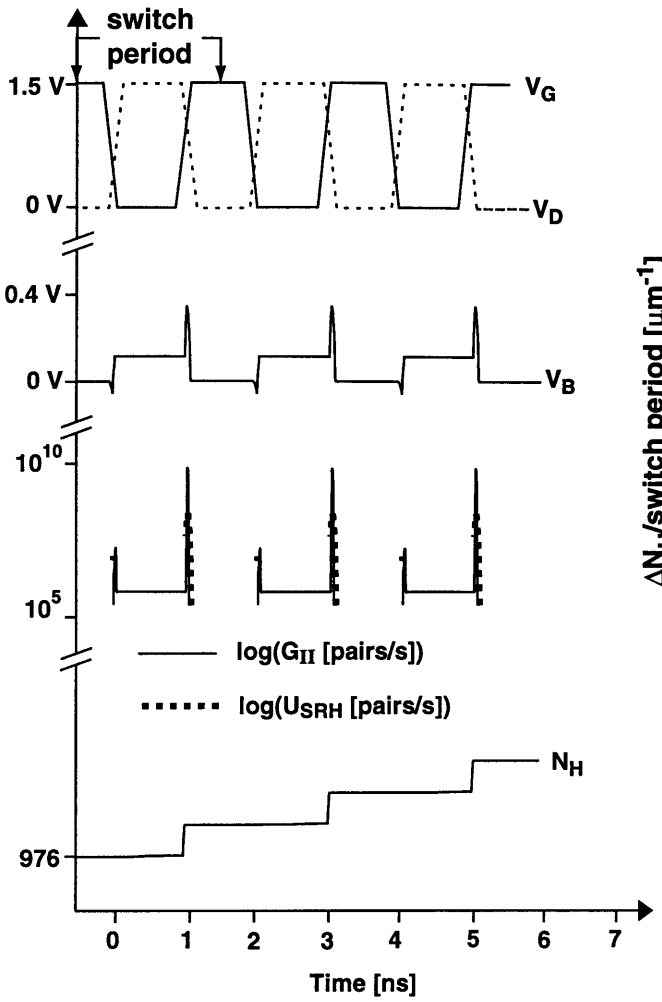
Figure 3-15 (b) shows what happens to the same NMOSFET if it starts OUT-LO. Starting OUT-LO, the initial body voltage is zero. It switches to OUT-HI, but to a $V_{B,pre}$ less than if it starts OUT-HI. As shown in Figure 3-15 (b), the body hole content increases with switching during the transition from OUT-HI to OUT-LO due to impact ionization hole generation being greater than diode hole recombination. This continues until a switching-steady-state is reached, the same



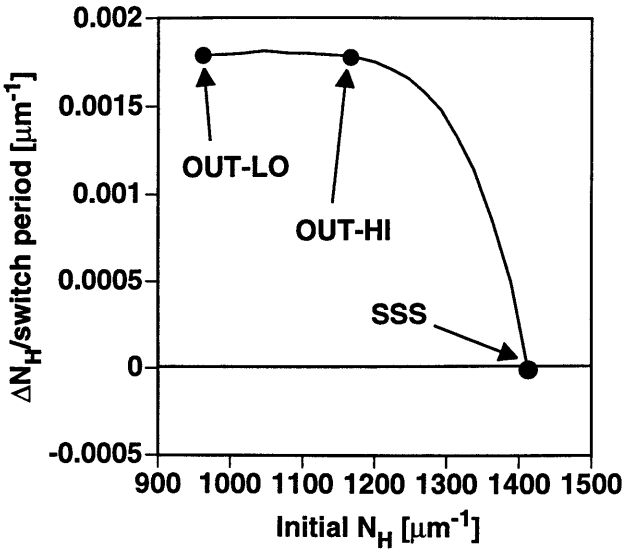
(a)

Figure 3-15. MEDICI-simulated example of PD-SOI NMOSFET (L_{eff}=0.12 μm, N_A=6x10^{17} cm^{-3}, N_D=1x10^{19} cm^{-3}, t_{ox}=4.5 nm, t_{si}=70 nm, t_{box}=200 nm) switching in CMOS inverter. (a) Switching starting from OUT-HI logic state. Shown are gate and drain terminal voltage waveforms, body voltage, impact ionization generation rate, hole recombination rate, and body hole content versus time. (b) Switching starting from OUT-LO. (c) Change in body hole content per switch period versus initial body hole content for the device used in this example. Note that the device is not charge-balanced so OUT-HI and OUT-LO initial conditions begin with different initial body hole content. Switching-steady-state is reached when the change in body hole content per switch period equals zero.

switching steady as starting OUT-HI. Figure 3-15 (c) summarizes the floating-body hole content behavior of this NMOSFET switching with the waveform pattern under discussion. Figure 3-15 (c) shows the change in body hole content over a switch period versus the initial body hole content at the beginning of the switch period. It shows that there are more holes in the body starting OUT-



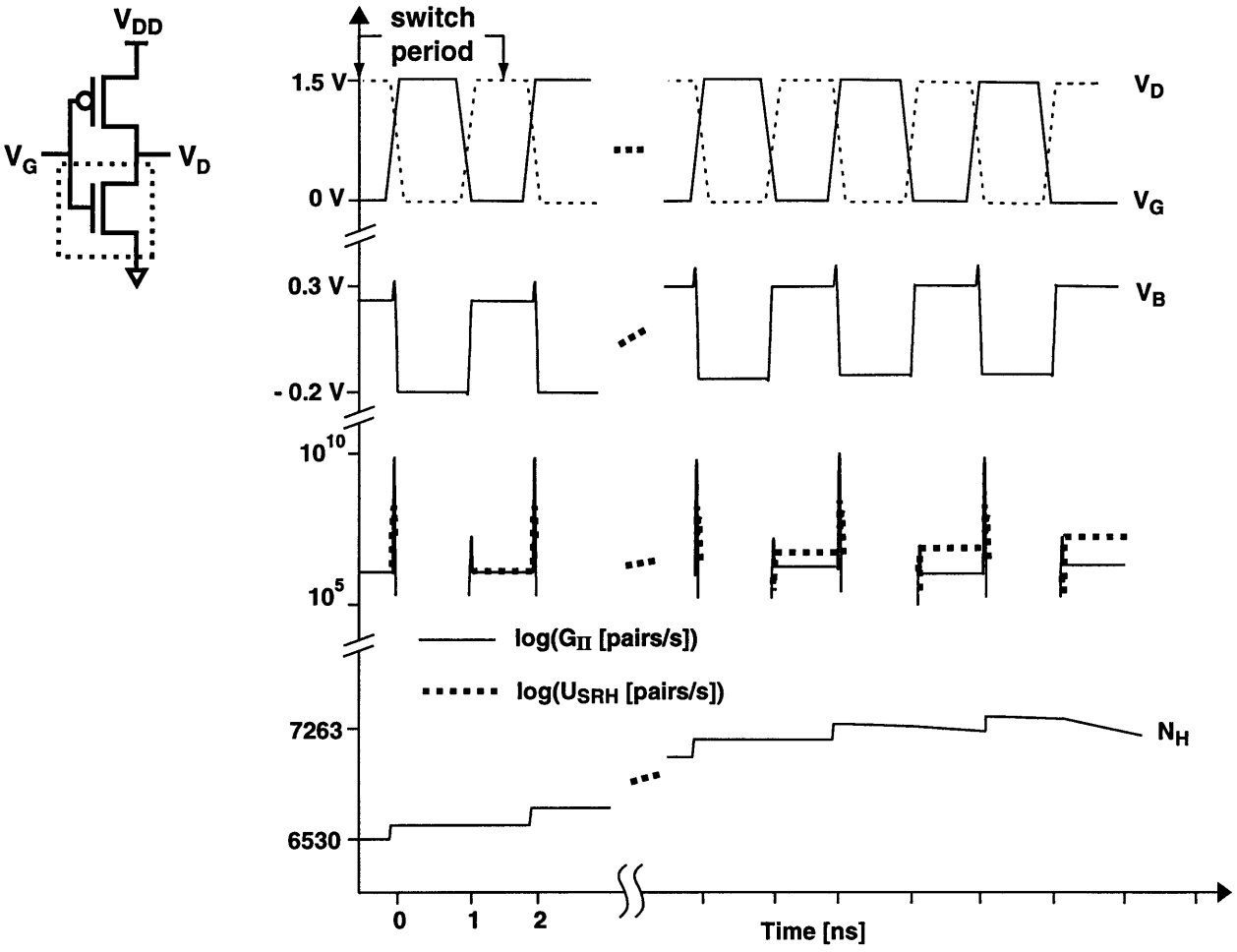
(b)



(c)

HI than starting OUT-LO, hence when starting OUT-LO and switching to OUT-HI, $V_{B,pre}$ is less than $V_{B,pre}$ starting OUT-HI. The curve shows that starting OUT-HI and starting OUT-LO, the body hole content increases over every switch period until the body hole content is sufficiently high to reach switching-steady-state, where there is net zero change in body hole content over one switching period. This increase in body hole content means that the NMOSFET has decreased threshold voltage in switching-steady-state. It now switches faster than it did starting OUT-HI, and even faster than it did starting OUT-LO.

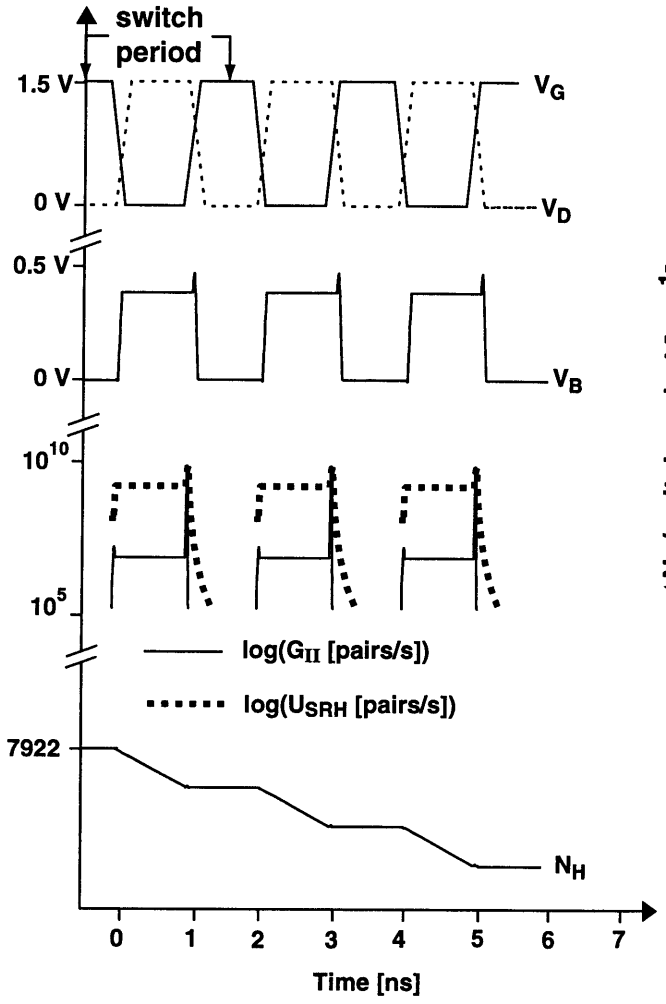
Figure 3-16 shows the floating-body behavior for a NMOSFET which is identical to that in Figure 3-15, except for a thicker silicon film. The thicker silicon film results in larger body-source/



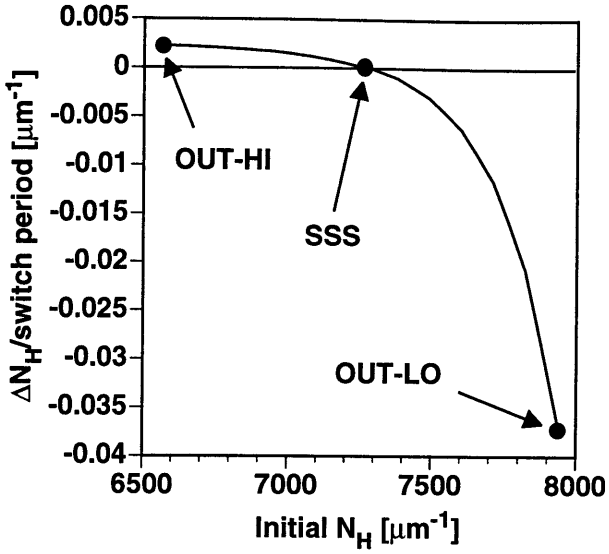
(a)

Figure 3-16. MEDICI-simulated example of PD-SOI NMOSFET (L_{eff}=0.12 μm, N_A=6x10^{17} cm^{-3}, N_D=1x10^{19} cm^{-3}, t_{ox}=4.5 nm, t_{si}=210 nm, t_{box}=200 nm) switching in CMOS inverter. (a) Switching starting from OUT-HI logic state. Shown are gate and drain terminal voltage waveforms, body voltage, impact ionization generation rate, hole recombination rate, and body hole content versus time. (b) Switching starting from OUT-LO. (c) Change in body hole content per switch period versus initial body hole content for the device used in this example. Only the silicon film thickness is different between this example and the example in figure 3-14. Note the different switching-steady-state behavior.

drain capacitance, which reduces body voltage coupling to the gate voltage, and increases body voltage coupling to the drain voltage. Starting OUT-HI, as shown in Figure 3-16 (a), the body voltage is pulled up by the gate much less than the thinner film device shown in Figure 3-15 (a). Impact ionization generation is not much different than in 3-15 (a), but recombination is much



(b)



(c)

reduced. Thus, the body hole content increases during the transition from OUT-HI to OUT-LO. When the drain voltage falls, the body voltage is pulled down to -0.2 V. Thus, the body-source/drain diodes are reverse biased in OUT-LO resulting in further generation of holes into the body. This is negative net recombination and thus does not appear on the plot of $\log(U_{SRH})$. The body hole content continues to increase during each switching period until the body voltage is sufficiently high that forward bias of the body-source/drain diodes over one switching period is sufficient to remove the holes generated by impact ionization and thermal generation, thus establishing switching-steady-state.

This device behaves quite differently starting OUT-LO as shown in Figure 3-16 (b). The

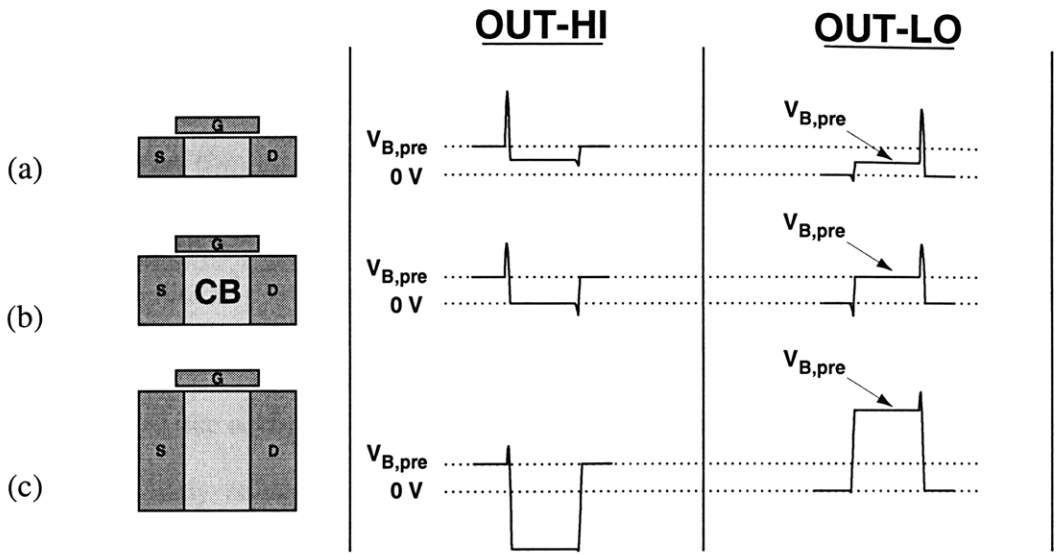


Figure 3-17. Schematic showing NMOS body voltage behavior during the first switching period starting from OUT-HI and OUT-LO for (a) silicon film thickness less than charge-balanced thickness, (b) charge-balanced, and (c) silicon film thickness greater than charge-balanced thickness. Please refer to figure 3-16 (a) and (b) for definition of the switching period starting from OUT-HI and OUT-LO, respectively. Note the charge-balanced body voltage waveforms are the same whether OUT-HI or OUT-LO, just offset by one half period.

body voltage is initially 0 V starting OUT-LO. When the NMOSFET switches to OUT-HI, the body voltage is pulled up by the drain voltage to a high value of $V_{B,pre}$. This high body voltage forward biases the body-source junction and removes holes form the body. This removal increases during the transition from OUT-HI to OUT-LO as the body voltage is pulled higher due to gate-body capacitive coupling. Recall that diode current is exponential with applied voltage so because $V_{B,pre}$ is already so high, the peak body voltage during the OUT-HI to OUT-LO transition is sufficiently high to remove all the holes generated by impact ionization during this transition. Thus, there is net loss of holes over one switching period, which continues until the body hole content and body voltage decreases such that hole generation and hole recombination are equal over a switch period. Again, it may take many thousands of switches to establish this switching-steady-state.

Figure 3-16 (c) summarizes the floating-body behavior of this device. Note the difference in behavior compared to Figure 3-15 (c). In this case, there are more holes in the body starting in OUT-LO than starting OUT-HI, and switching steady state is somewhere in between. This means

that a device starting OUT-HI will get faster while switching, and a device starting OUT-LO will slow down while switching. Recall that the only difference between the devices in Figure 3-15 and 3-16 is the silicon film thickness.

Switching behavior can be summarized with schematics of body voltage behavior versus silicon film thickness shown in Figure 3-17. The body voltage is shown over one switch period starting OUT-HI and starting OUT-LO for three MOSFET's. The first device has t_{si} less than charge-balanced, the second is charge-balanced, and the third has t_{si} greater than charge-balanced. The preswitch body voltage is greater starting OUT-HI than starting OUT-LO for the thinner film device, while it is greater starting OUT-LO than starting OUT-HI for the thicker film device. This indicates a ΔV_T between starting OUT-HI and starting OUT-LO. Over many switching periods, the OUT-HI and OUT-LO waveforms either increase or decrease in overall levels, and will both converge to a switching-steady-state waveform. Notice that when the charge-balanced device has the same preswitch body voltage starting OUT-HI and OUT-LO. In fact, the OUT-HI and OUT-LO body voltage waveforms are identical, just offset by a half switch period. This indicates zero ΔV_T between starting OUT-HI and starting OUT-LO. Over many switching periods the body voltage levels will either increase or decrease until switching-steady-state, but ΔV_T between logic states remains zero. This is minimizes threshold variation and eliminates pulse stretching, as shown in Figure 3-9.

3.7 Simulation Methodology to Determine Threshold Voltage Variation

The 2-D numerical simulator MEDICI can be used to determine the threshold voltage starting OUT-HI, starting OUT-LO, and at switching-steady-state. In each case, transient simulations are used to produce the instantaneous gate transfer *I-V* characteristics from which V_T can be extracted. The minimum set of models used are what are required to describe a MOSFET, diodes, and impact ionization generation. These include low-field, perpendicular-field, and high-field mobility models,

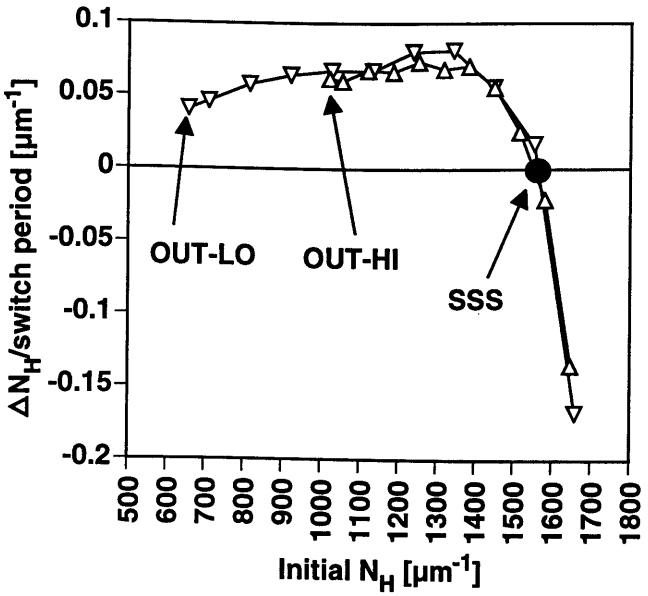


Figure 3-18. MEDICI simulation example of finding switching-steady-state body hole content for a NMOSFET. Each symbol represents the same simulation structure at a different initial body hole content, run through a switching cycle. Each curve represents simulations where the switch cycle starts at OUT-HI or OUT-LO. The body hole content which yields zero change in body hole content after a switching cycle is the switching-steady-state body hole content.

and 2-carrier solution, Shockley-Read-Hall generation-recombination statistics, Auger recombination, bandgap narrowing, and impact ionization [37].

Simulation for threshold voltage starting OUT-HI is the most straightforward. The first step is to solve for a DC solution with V_G=V_S=0 V and V_D=V_{DD}. It is not possible to simply ramp the gate voltage because the displacement current and the drain current cannot be easily separated. Instead, multiple simulations quickly ramp the gate to a particular voltage. For example, to trace out I_{DS}-V_{GS} with data points at 0.1 V gate voltage intervals up to V_{GS}=0.7 V, seven separate transient simulations are done. The first ramps the gate to 0.1 V, the second ramps the gate to 0.2 V, etc. This is done with a ramp rise time of 10 ps. This is short enough time for the drain current to be the instantaneous value. This risetime can be shorter if necessary. Once the gate reaches the end of the ramp, the displacement current settles within one or two timesteps leaving only the drain current. The results for various V_G ramps are compiled, plotted, and V_T is extracted at a constant current.

Simulation for threshold voltage starting OUT-LO has one added step. First, solve for a DC

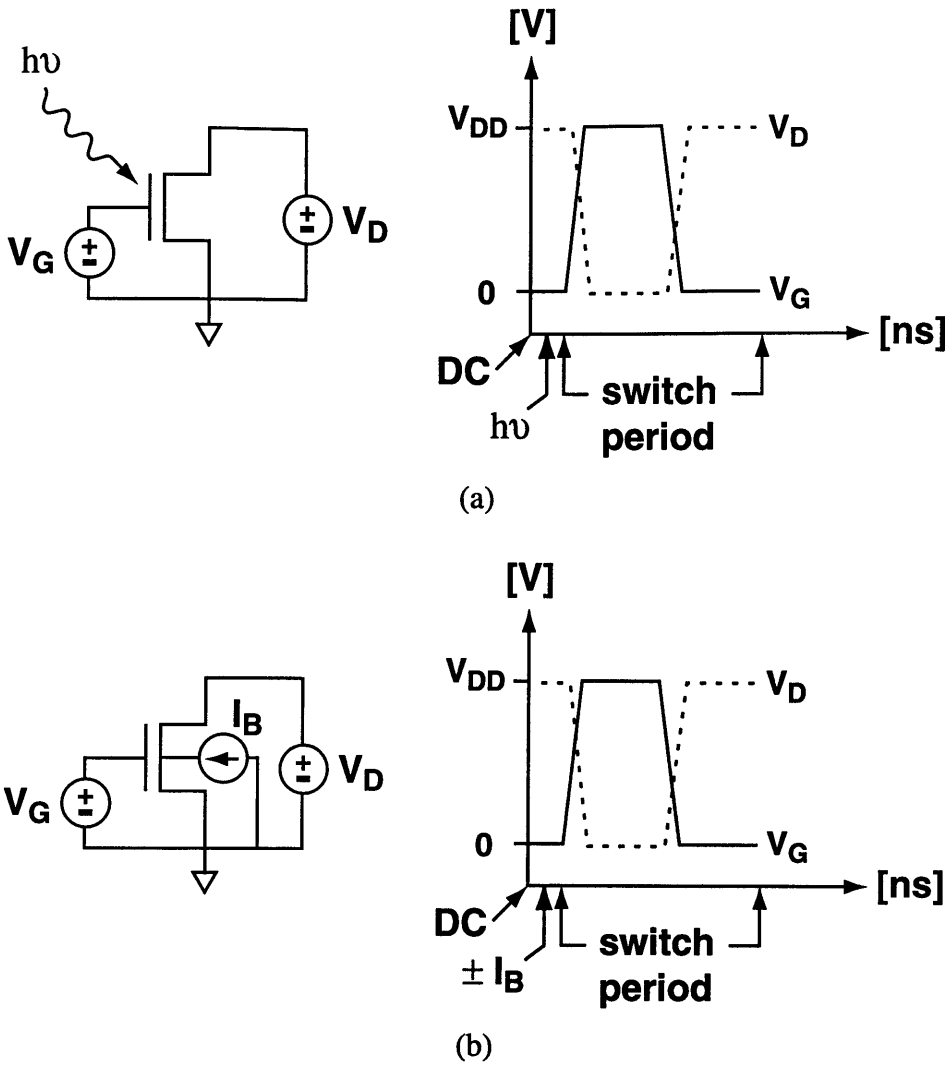


Figure 3-19. MEDICI simulation methodology for finding switching-steady-state. The MOSFET is simulated with external SPICE elements used to mimic NMOS switching in a CMOS inverter. The waveforms are shown to the right of the circuit. Initial body hole content is set prior to the switching cycle by (a) photogeneration into a floating body or (b) an ideal current source into a contacted body.

solution with V_D=V_S=0 V and V_G=V_{DD}. This NMOSFET is in OUT-LO. It must be switched to OUT-HI. This is done with a rapid ramp of V_G to 0 V, then a rapid ramp of V_D up to V_{DD}. Note that it is important to keep the order of terminal voltage ramps because pulling the drain voltage up with the gate voltage still up may cause significant changes in body hole content due to the body voltage being pulled up enough to strongly forward bias the body-source diode, and due to significant impact ionization generation because the device is turning ON. Once switched to OUT-HI, the

same procedure used for starting OUT-HI is used to extract V_T.

Simulation for the threshold voltage of a device in switching-steady-state is not that much more difficult. First, the switching-steady-state body hole content must be found. It is impractical to simply run a transient simulation with the switching waveform pattern over many thousands of switching periods. Thus, the following method is used. Starting with a OUT-HI or OUT-LO DC solution, the body hole content in this solution is changed, as disucssed below, and the body hole content is integrated. This is the initial body hole content before switching. The simulated device is then subjected to one period of the switching waveform, after which the body hole content is again integrated. A plot is made of the change in body hole content per switch period versus the initial holes in the body. The initial body hole content which results in zero change in body hole content over a switch period is the switching-steady-state body hole content. An example is shown in Figure 3-18. Curves are shown for a starting OUT-HI DC solution and a starting OUT-LO DC solution. Both solutions end up at a switching-steady-state solution which is unique to the switching waveform. Once this solution is found, a starting OUT-HI DC solution has it's body hole content modified to the switching-steady-state body hole content and the same procedure used starting OUT-HI is used to extract V_T.

There are two methods to modify the body hole content of a DC solution. The first is shown in Figure 3-19 (a) and involves a photogeneration module in MEDICI which allows for either electrons or holes to be generated versus time. The generated carriers incorporate themselves into the solution on the order of the dielectric relaxation time allowing the device to be in equilibrium prior to application of the switching waveforms.

The one disadvantage of photogeneration is that while it allows for increase in body majority carrier content, it does not allow for decrease. The second method allows for both increase and decrease in body majority carrier content. It involves use of the Circuit Advanced Application Module (CAAM) in MEDICI which allows ideal SPICE elements to be connected to a 2-D numerical simulation mesh for "mixed-mode" simulation. To mimic a floating-body and allow the body majority carrier content to be set, a body-contacted version of the simulation structure is created

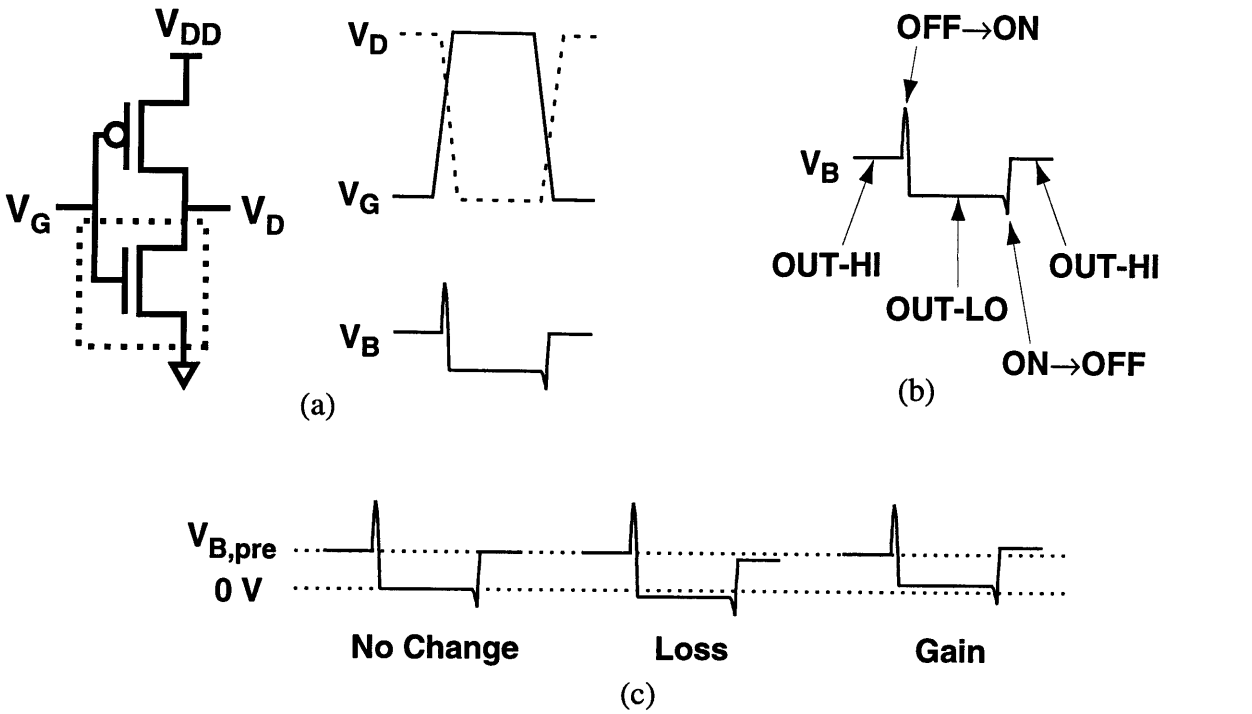


Figure 3-20. (a) Switching period definition starting from OUT-HI and body voltage response. (b) Body voltage relationship to MOSFET logic state and transitions between logic states OUT-HI and OUT-LO. (c) Schematic showing charge-balanced body voltage response to one switching period in which there is no change, a loss of body hole content, or a gain in body hole content (loss and gain are not to scale).

with an ideal current source at the body contact, as shown in Figure 3-19 (b). Positive current is used to add holes and negative current is used to remove holes form the DC solution.

3.8 Minimizing Threshold Voltage Variation

As has been shown, designing the silicon film thickness for charge-balance minimizes threshold voltage variation because the two logic states OUT-HI and OUT-LO have the same equilibrium body hole content. However, the body hole content is out of equilibrium during transitions between OUT-HI and OUT-LO, and can change during these transitions. Figure 3-20 shows schematically the change in preswitch body voltage for no change, loss of holes, or a gain in holes through one switch period. The majority of the change occurs during the OFF to ON transition shown in Figure 3-20 (b) when impact ionization generation tends to increase body hole content

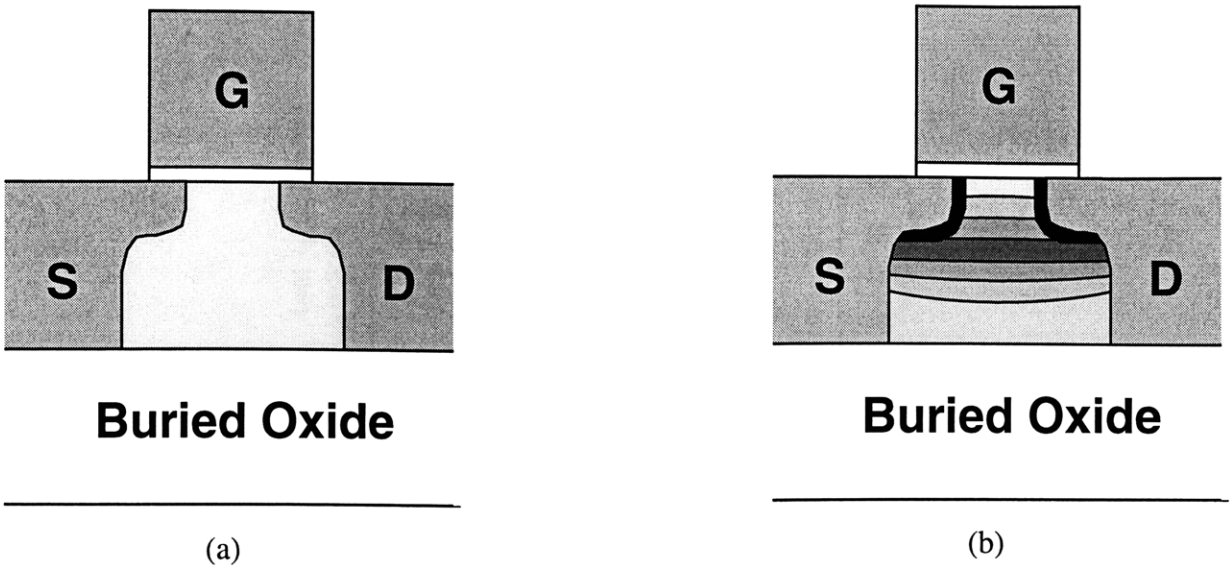


Figure 3-21. Schematic cross section of SOI MOSFET's with (a) uniform body doping and (b) nonuniform body doping retrograde channel doping and source/drain halos. The nonuniformly doped body has a higher peak doping resulting in higher diode ideality factor.

and the body-source diode conduction tends to decrease body hole content.

It is impractical to try to balance the precise number holes generated by impact ionization with the number recombined by diode conduction through the body-source diode. This is because switching loads vary in size and therefore the precise waveforms for which this balance is achieved is not common to every device. A better way to address both NMOS and PMOS threshold voltage variation is to raise diode current levels so that they make impact ionization current less significant in NMOSFET's. Also, making the forward diode and reverse diode currents more symmetric should allow for any loss of holes through forward biased diodes to be more easily recovered from the reverse biased diodes. And probably the best way to minimize threshold voltage variation in both NMOS and PMOS is to reduce to power supply voltage. Reduction of power supply voltage reduces TURN-ON impact ionization and reduces capacitive coupling of the body to the gate voltage during TURN-ON, minimizing both gain and loss of body majority carriers.

To demonstrate these concepts, the threshold voltage variation in NMOSFET's and PMOSFET's is simulated in MEDICI. A silicon film thickness which is closest to charge-balancing both NMOS and PMOS is chosen in each case. Uniform and nonuniform body doping profiles are

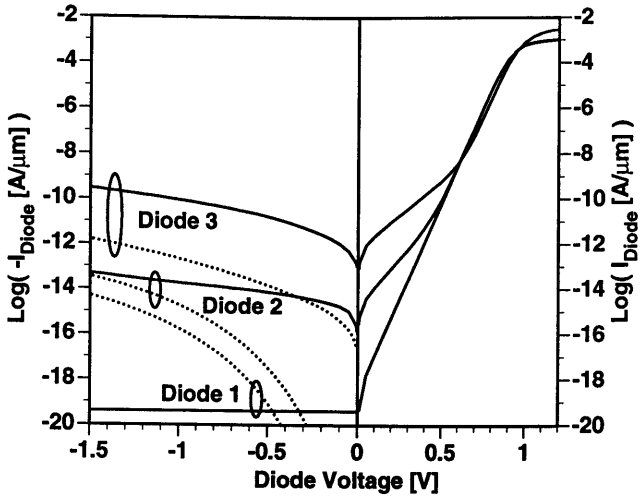


Figure 3-22. Sample body-source/drain diode *I-V* characteristics simulated in MEDICI used as reference for figure 3-23. The labels refer to forward diode ideality factor.

compared, as shown schematically in Figure 3-21. The body doping was designed so that the DC threshold voltage at V_{DS}=100 mV is 0.35 V. The uniformly doped devices have diode *I-V* characteristics which look like the diode 2 in Figure 3-22. The nonuniformly doped devices have higher peak doping and thus look more like diode 3, with higher current levels and more diode symmetry. Note that in both cases the reverse bias diode current is greater than NMOS OFF-state impact ionization.

Figure 3-23 (a) shows the threshold voltage variation for devices designed with 0.18 μm technology dimensions (L_{GATE}=0.18 μm, L_{eff}=0.12 μm, t_{ox}=2.5 nm, x_j=60 nm) at a power supply of 1.5 V. The threshold voltage starting OUT-HI and starting OUT-LO is equal due to charge-balance. The switching-steady-state threshold voltage reflects an applied waveform of a CMOS inverter switching between OUT-HI and OUT-LO at 50 % duty cycle over a period of 2 ns. The fall and rise times of the gate and drain voltage waveforms is 50 ps, with a 40 ps delay between the gate and drain transitions. The nonuniformly doped NMOS has smaller V_T variation than the uniformly doped NMOS because of higher diode current levels keeping impact ionization in check. The PMOS V_T variation is only slightly smaller for the nonuniformly doped devices indicating small contribution from diode symmetry. In fact, most of the difference may be due to the larger body-source/drain capacitance of the nonuniformly doped devices keeping the body voltage from increasing

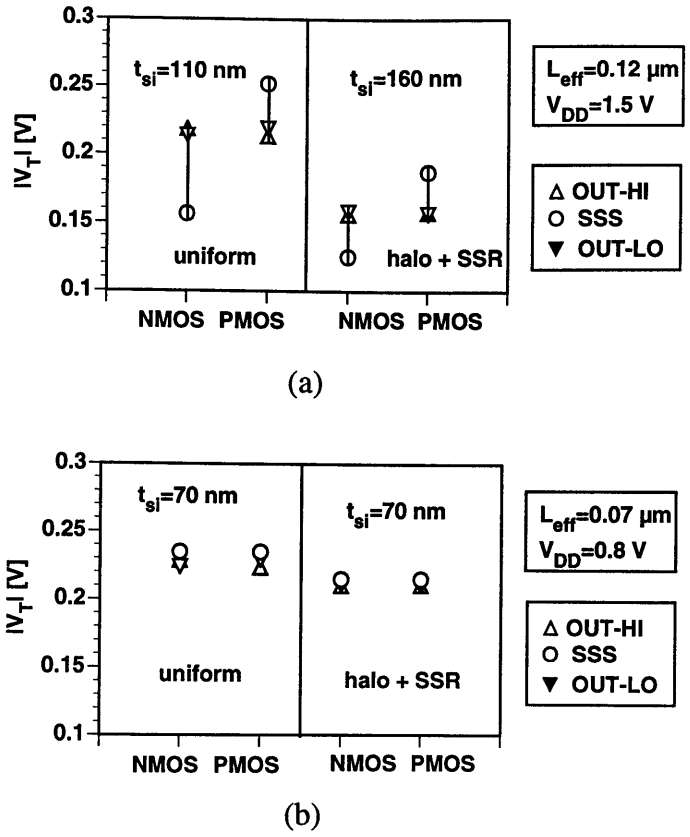


Figure 3-23. MEDICI simulated threshold voltage hysteresis for different doping profile types for (a) 0.18 μm technology and (b) 0.13 μm technology designs. The nonuniformly doped halo+SSR profiles show reduced threshold voltage hysteresis compared to uniformly doped profiles. Lower supply voltage dramatically reduces threshold voltage hysteresis.

as much as in the uniformly doped devices. This limits the loss of majority carriers through forward diode bias. Had the switching waveform had time between switching transitions, increased diode symmetry would be more important as there would be sufficient time between switches for reverse-biased diodes to help replenish the deficit of majority carriers.

Figure 3-23 (b) shows the threshold voltage variation for devices designed with 0.13 μm technology dimensions (L_{GATE}=0.13 μm, L_{eff}=0.07 μm, t_{ox}=2 nm, x_j=40 nm) at a power supply of 0.8 V. The reduced power supply greatly reduces threshold voltage variation. Again, the nonuniformly doped devices have reduced threshold voltage variation. This is due to a combination of higher body-source/drain capacitance keeping the body from being pulled up by the gate and perhaps more symmetric diodes.

Overall, MOSFET scaling tends to help reduce hysteretic threshold voltage variation. Higher doping levels are required for electrostatic integrity leading to increased diode current levels, diode symmetry, and increased body-source/drain capacitance. The latter along with reduced power supply decreases impact ionization and decreases changes in body voltage during TURN-ON responsible for loss of body majority carrier content. Reduced power supply also decreases impact ionization. Note however that reduced body voltage coupling to the gate decreases current boost and degrades performance, as does reducing power supply. Ultimately, silicon film thickness design for charge-balance is still the most beneficial method to reduce threshold voltage variation induced pulse stretching and like effects. Without charge-balance, the threshold voltage variation in Figure 3-23 (b) can be substantially more than is shown.

3.9 Summary

This chapter showed hysteretic behavior of floating-body majority carrier content in PD-SOI MOSFET's. This floating-body-induced hysteresis leads to threshold voltage variation which becomes very important to circuit operation as power supplies are reduced. This was demonstrated in measurements of threshold voltage variation leading to stretching of the input pulse length as it propagated through a CMOS inverter chain. A simple guideline for designing the silicon film thickness for "charge-balance", that is, equal majority carrier content in equilibrium regardless of logic state was demonstrated to be effective in reduction of pulse stretching.

Figure 3-24 summarizes the conclusions made in both this chapter and the previous chapter. It does so by showing pulse propagation trends in CMOS inverter chains where the MOSFET's have silicon films thinner than charge-balanced, and thicker than charge-balanced. When thinner than charge-balanced, the first pulse stretches as it propagates. In worst-case switching-steady-state, the pulse width is recovered but the rising edge slows down more than the falling edge speeds up, due to a greater tendency for devices to slow down rather than speed up. This is due to the asymmetry between forward and reverse diode characteristics which make it easier for the body to lose majority carriers than gain majority carriers. Impact ionization contributes to gain of body

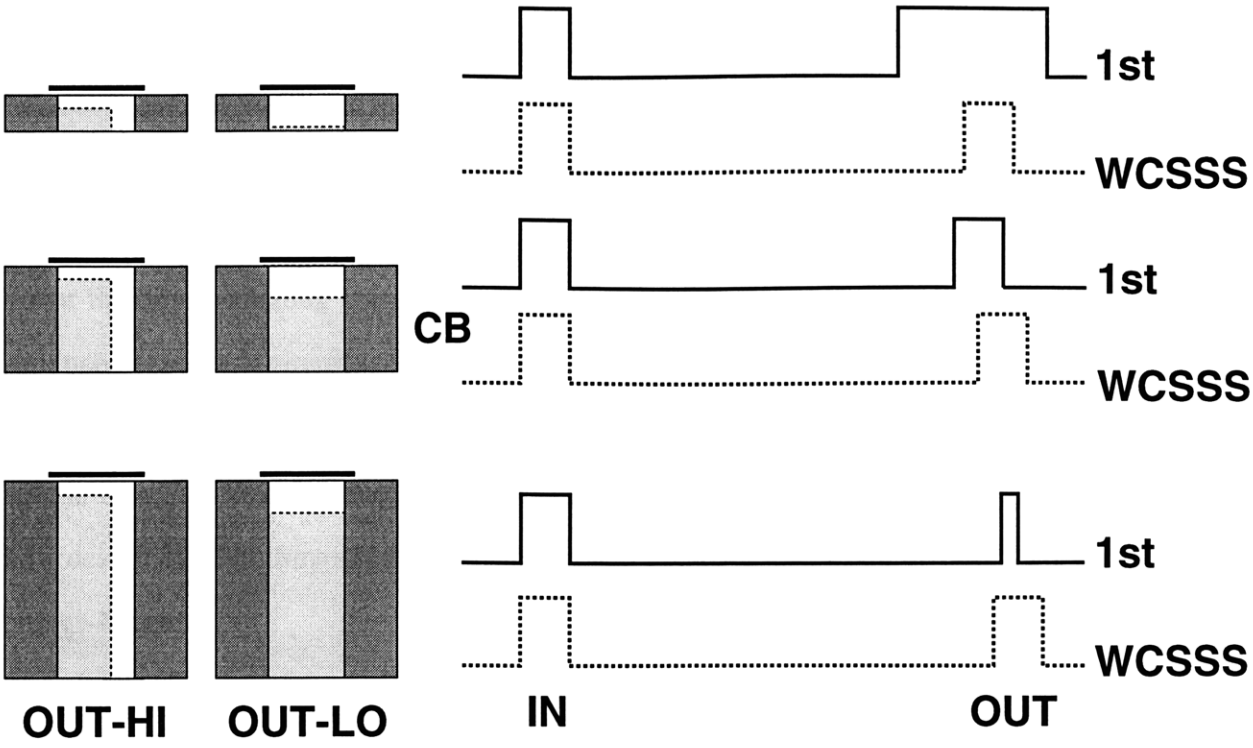


Figure 3-24. Schematic showing pulse propagation through a CMOS inverter chain versus silicon film thickness relative to charge-balance. (a) When the silicon film thickness is less than charge-balanced, the first pulse through the inverter chain stretches. In worst-case switching-steady-state the output pulse recovers the input pulsewidth, but the midpoint of the pulse has slowed down relative to the first pulse. (b) When charge-balanced, the output pulse is always the same width as the input pulse, but slows down once it reaches worst-case switching-steady-state. (c) When the silicon film thickness is greater than charge-balanced, the first pulse through the inverter chain compresses, provided that the inverter chain is not long enough for the pulse to disappear altogether. In worst-case switching-steady-state the output pulse recovers the input pulsewidth, but the midpoint of the pulse has slowed down relative to the first pulse.

majority carriers, particularly in NMOSFET's, but no published measured pulse stretching data has shown a definitive speed up due to impact ionization [12-18]. For charge-balanced devices, the pulse does not stretch, but pulse propagation should slow down at worst-case switching-steady-state. And finally, for thicker than charge-balanced, the first pulse compresses. The pulsewidth is recovered at worst-case switching-steady-state, but rising edge should tend to speed up less than the falling edge slows down, again due the tendency for devices to lose body majority carriers.

The worst-case switching-steady-state propagation delay for the thinner than charge-balanced

case is shown to be faster than for charge-balanced, and even faster than the thicker than charge-balanced case. As was noted in chapter 2, the thinner silicon films should produce faster MOSFET's due to reduced body-source/drain capacitance allowing more gate-body pull-up of the body voltage thereby boosting the drain current, and less drain-body pull-down of body voltage which loads down the next switching transition. It can be concluded that overall, thinner silicon films can produce increases in circuit speed, as shown in both the first pulse and worst-case switching-steady-state. But this comes at the expense of hysteresis in propagation delays. Thus, designing the silicon film thickness for charge-balance is perhaps the best strategy. The next chapter will show how design of other dimensions can be optimized to maximize switching speed while maintaining charge-balance.

Chapter 4

PD-SOI CMOS Technology Optimization

The behavior of partially-depleted SOI MOSFET's with the body floating has been discussed in the previous chapters. Understanding of the floating body in a PD-SOI MOSFET allows design which exploits its advantages and minimizes its drawbacks. In this chapter, floating-body SOI CMOS technology, which is optimized to maximize drive current while minimizing threshold voltage hysteresis, is compared to bulk CMOS technology using the 2-D numerical simulator MEDICI. As before, the focus is on NMOSFET's, but recall (from section 3.4) that PD-SOI PMOSFET design is similar,particularly if OFF-state impact ionization is mitigated by reverse diode leakage.

4.1 OFF-Current versus ON-Current Figure of Merit

Of all the MOSFET parameters discussed in chapter 2, the most important are the OFF-current, I_{OFF}, and the saturated ON-current, I_{ON}. I_{OFF} and I_{ON} are defined in Figure 4-1 (a). At a given gate length, a MOSFET built with a CMOS technology is said to be superior to a MOSFET built with another CMOS technology if it possesses higher I_{ON} for the same I_{OFF}, lower I_{OFF} for the same I_{ON}, or higher I_{ON} and lower I_{OFF}.

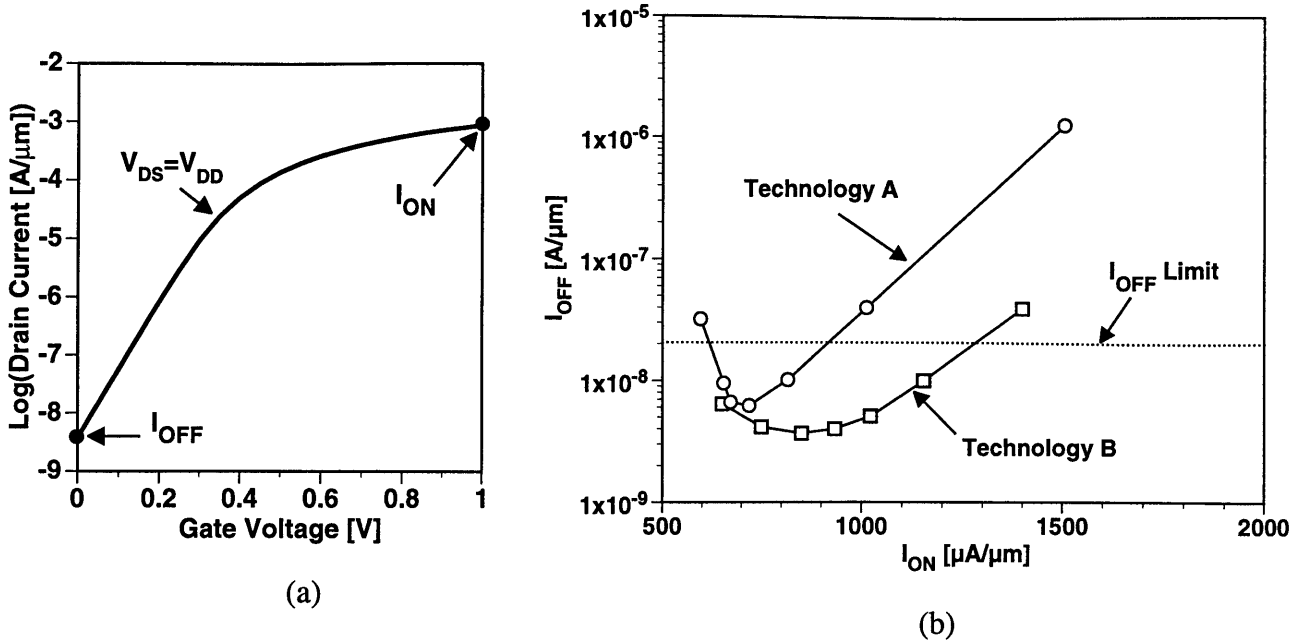


Figure 4-1. (a) Definition of OFF-current and ON-current on a gate transfer *I-V* characteristic. (b) Log OFF-current versus ON-current example for two hypothetical technologies. From left to right, each symbol on each curve represents a gate length. For example, the first symbol on the left of each curve represents $L_{GATE}=L_1$, the next on each curve represents $L_{GATE}=L_2$, etc.

In this context, "CMOS technology" needs more definition. A CMOS technology is characterized by a set of intrinsic device design variables, such as power supply voltage (which determines the maximum internal electric fields), gate oxide thickness, source/drain junction depth, doping profiles, OFF-current limit, etc. The physical size of the transistor, and hence the size of extrinsic parasitic capacitances, is characterized by layout design rules based on a feature size [38]. A figure of merit for a CMOS technology is to plot I_{OFF} on a logarithmic scale versus I_{ON} on a linear scale, for various gate lengths. Due to variations in manufacturing, the gate length varies around a nominal value. Thus, the I_{OFF} versus I_{ON} curve shows this variation of I_{ON} and I_{OFF} around a nominal gate length. The better the CMOS technology, the more it minimizes I_{OFF}, maximizes I_{ON}, while minimizing the slope of log I_{OFF} versus I_{ON}.

An example is shown in Figure 4-1 (b), which compares two hypothetical CMOS technologies. Each symbol on each technology curve corresponds to the same physical gate length, L_1, L_2, L_3, etc., from left to right. Technology B is superior because its log I_{OFF} to I_{ON} curve is lower, farther right, and has a shallower slope than the curve representing Technology A. At each gate

length, Technology B devices have less I_{OFF}. Technology B I_{ON} is not always greater than Technology A, but is greater at each gate length for devices below the I_{OFF}-limit, above which devices cannot be used anyway. The shallower slope of the right part of the Technology B curve is indicative of better electrostatic integrity, since I_{OFF} degrades less with shrinking gate length.

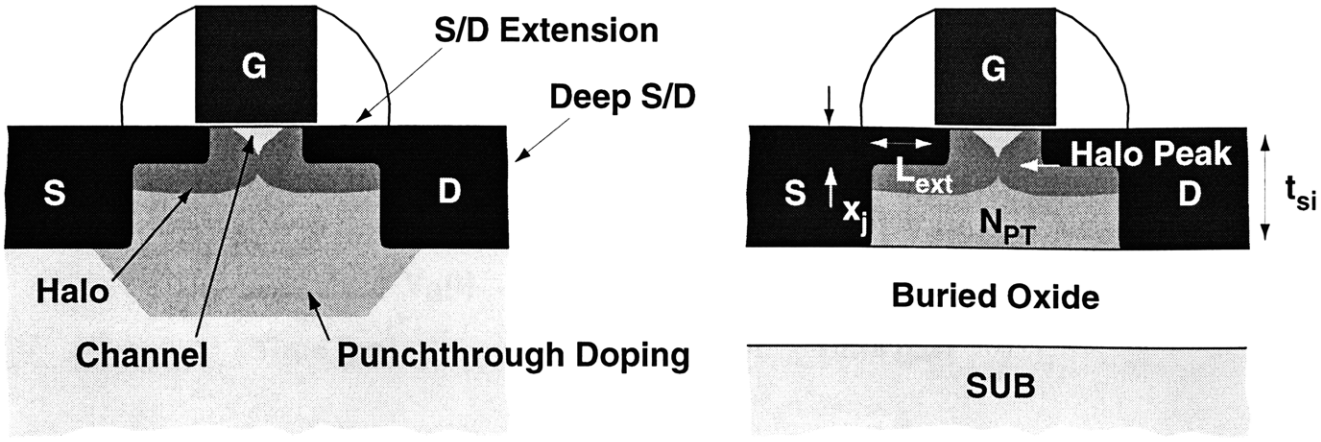
The "hook" shown on the left side of each curve is a result of the reverse short-channel effect. This typically occurs in short-channel MOSFET's designed with source/drain halos. As the channel length decreases, the halo dopings overlap, increasing the channel doping, thereby decreasing I_{OFF} with decreasing gate length. Less "hook" is desirable since this results in less variation of I_{OFF} with gate length. Figure 4-1 (b) shows that Technology B has less "hook" than Technology A.

4.2 Design Variables and Optimization Framework

Optimization of the intrinsic bulk MOSFET is fairly straightforward. The International Technology Roadmap for Semiconductors publishes guidelines for the gate oxide thickness, junction depth, I_{OFF} limit, and threshold voltage variation with gate length, for given ranges of gate lengths grouped into a technology [28]. Selected 100 nm-, 70 nm-, and 50 nm-technology guidelines which will be used in this study are listed in Table 4-1. These guidelines include the maximum, nominal, and minimum gate lengths for the technology (L_{MAX}, L_{NOM}, L_{MIN}), gate-source/drain overlap

Technology	100 nm	70 nm	50 nm
L_{MAX}	78 nm	54 nm	38 nm
L_{NOM}	65 nm	45 nm	32 nm
L_{MIN}	52 nm	36 nm	26 nm
ΔL	21 nm	15 nm	10 nm
V_{DD}	1.0 V	0.9 V	0.7 V
$t_{ox,eq}$	1.25 nm	1 nm	0.7 nm
x_j	32.5 nm	24 nm	16.5 nm
$I_{OFF}@L_{MIN}$	20 nA/μm	40 nA/μm	80 nA/μm
ΔV_T(NOM-MIN)	33 mV	25 mV	17 mV

Table 4-1. International Technology Roadmap for Semiconductors criteria used to create doping profiles.



(a) (b)

Figure 4-2. (a) Definition of terms in scaled MOSFET's. Diagram shows a bulk MOSFET. (b) Definition of design variables in MOSFET's. Diagram shows a SOI MOSFET. L_{ext} is the source/drain extension length (or the spacer width), N_{PT} is the punchthrough doping density, and t_{si} is the silicon film thickness.

(ΔL), power supply voltage (V_{DD}), equivalent gate oxide thickness ($t_{ox,eq}$), source/drain extension junction depth (x_j), I_{OFF}-limit for the minimum gate length device, and threshold voltage variation between the nominal and minimum gate length devices. With these guidelines in place, only the doping profiles remain to be optimized. As shown in Figure 4-2 (a), these doping profiles consist of the channel doping, source/drain extension doping, deep source/drain doping, and the punchthrough doping. Of these, only the channel and source/drain extension doping affect the intrinsic bulk MOSFET. The deep source/drain only serves to allow good contact formation to the source/drain extensions, and the punchthrough doping only serves to eliminate leakage from the deep source to the deep drain.

The optimized bulk MOSFET can be directly transferred to SOI. However, because of the floating body, the electrostatics for SOI MOSFET's are modified by $V_{B,OFF}$, and I_{ON} is modified by the body voltage boost due to capacitive coupling from the gate to the body, as shown in Figure 4-3. Recall from chapter 2 that changes in the body-source/drain diodes affect the value of $V_{B,OFF}$. Also recall from chapter 2 that $V_{B,OFF}$ and the body-source/drain capacitance affect gate-body capacitive coupling. Because the body-source/drain diodes and capacitances depend on the

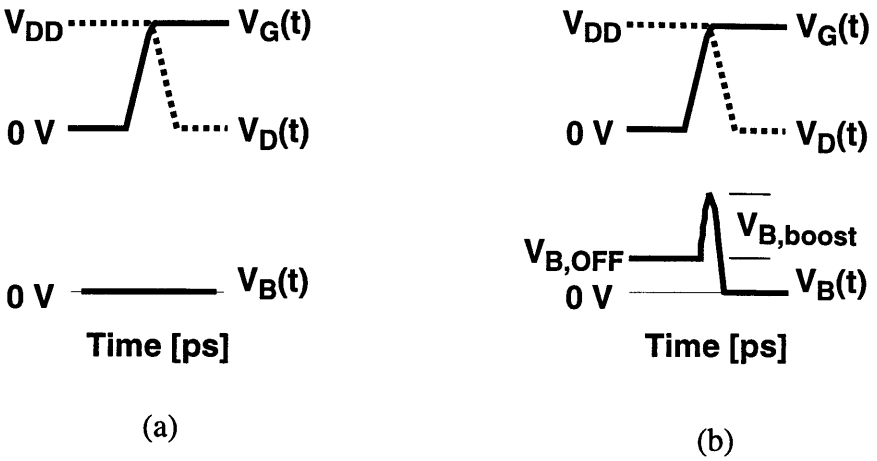


Figure 4-3. Comparison of body voltage during switching for (a) bulk NMOSFET's and (b) SOI NMOSFET's. The SOI NMOSFET has V_B=$V_{B,OFF}$ in the OFF-state, and turns ON with extra body voltage boost due to capacitive coupling.

dimensions and dopings of the entire body-source/drain junction, the deep source/drains, punchthrough doping, source/drain extension length, and the silicon film thickness all contribute to intrinsic SOI MOSFET behavior.

SOI MOSFET's will require further redesign for a fair comparison between bulk and SOI. I_{OFF} will increase in SOI MOSFET's due to the body voltage floating to $V_{B,OFF}$ in the OFF-state. I_{ON} increases because $V_{B,OFF}$ lowers the threshold voltage. This change in I_{OFF} and I_{ON} due to $V_{B,OFF}$ can be eliminated by reoptimizing the channel doping profile of the SOI MOSFET's so that I_{OFF} is matched to bulk MOSFET I_{OFF} at the nominal gate length.

From the discussion above, the following simulation strategy is used to compare bulk to SOI. Since only I_{OFF} and I_{ON} are being compared, the SOI simulations are done with the same models, structure, and grid as bulk, except for modified doping profiles to match I_{OFF} at the nominal gate length and modified body voltage to include $V_{B,OFF}$ and body voltage boost. This ensures that the simulation models, structure, and grid are consistent between bulk and SOI, which is important because variations in the simulation grid or use of different models can create quantitative differences even when simulating the exact same structure [37]. $V_{B,OFF}$, body voltage boost, and charge-balance are calculated in a separate SOI grid with models, structure, and grid spacing optimized for calculating diode tunneling currents, impact ionization, and integrating for body hole content (grid spacing

should not scale with dimensions because this leads to artifacts in diode tunneling and charge integration). This methodology requires the following steps. 1) Create a bulk MOSFET structure and optimize bulk doping profiles to meet specifications in Table 4-1. This can be done by using gaussian profiles for the source/drain extensions and halos, as shown in Figure 4-4. The depth of the halo doping can be varied. 2) Create a SOI MOSFET structure using the optimized bulk doping profile. Add deep source/drains and punchthrough doping. 3) Optimize the source/drain extension length, punchthrough doping, and silicon film thickness to maximize body voltage boost and minimize threshold voltage hysteresis through charge-balance. 4) Assess $V_{B,OFF}$ for the nominal gate length device. 5) Apply $V_{B,OFF}$ to nominal gate length bulk structure and modify the halo doping so that I_{OFF} matches I_{OFF} with V_B=0 V at the nominal gate length. 6) Re-optimize SOI MOSFET structure using modified doping profile and assess $V_{B,OFF}$ and body voltage boost for each gate length device.

At this point it is possible to calculate I_{OFF} and I_{ON} for bulk MOSFET's at various gate lengths using the original optimized doping profile from step 1. Then, using the modified doping profile from step 5 and $V_{B,OFF}$ from step 6, calculate I_{OFF} for SOI MOSFET's at various gate lengths. I_{ON} for SOI MOSFET's can be calculated for each gate length by adding the body voltage boost of each gate length to the simulation. The result is a fair comparison of I_{OFF} versus I_{ON} curves for bulk and SOI of the same technology and using the same simulation models, structure, and grid. The following sections detail the above steps and show simulation results.

4.3 Simulation Details

Figure 4-4 shows the left half of the simulation structure used to optimize source/drain extension and halo dopings. The doping profiles are characterized by a peak value and gaussian rolloff parameters s_x and s_y in the depthwise and lateral directions, respectively. An optimization routine is run which varies the gate length, peak halo doping, gaussian parameters for the halo, and gaussian parameters for the source/drain extension until the specifications of Table 4-1 are met.

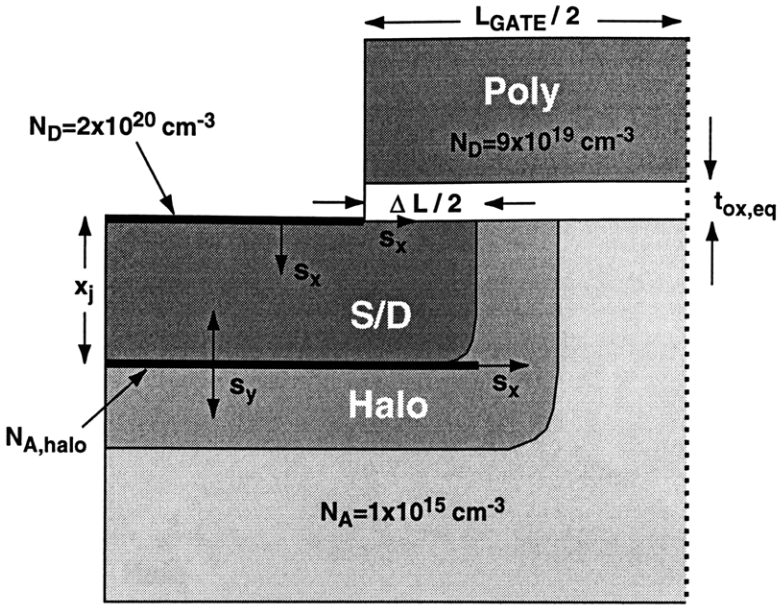


Figure 4-4. Left half of simulation structure used to optimize doping profiles to meet specifications in table 4-1. Doping profile parameterization variables are shown. The halo is to be optimized at the junction depth in this example.

Figure 4-5 shows an sample doping profile optimized with the halo at the source/drain extension junction depth.

Two sets of optimizations were done for each technology node in Table 4-1. The first optimized the doping profiles for the halo peak at the source/drain extension junction depth, x_j. The second optimized the doping profile for the halo peak at half the source/drain extension junction depth, $x_j/2$. Table 4-2 lists the resultant optimized doping profile parameters for the 100 nm-, 70 nm-, and 50 nm-technology nodes. Note that to simplify doping profile optimization, s_x for the source/drain extension profile is set equal to $\Delta L/3$.

NMOSFET I_{OFF} and I_{ON} were calculated in MEDICI with hydrodynamic modeling which simultaneously solved for potential, electron continuity, and electron temperature at each node. The models used are as follows: Fermi-Dirac statistics, band-gap narrowing, concentration dependent mobility for low-fields, a mobility model consistent with the universal mobility curve [39] for perpendicular fields, and an electron-temperature dependent mobility model was used for high lateral fields, with the energy relaxation time set at 0.2 ps. Poly-depletion is accounted for with a

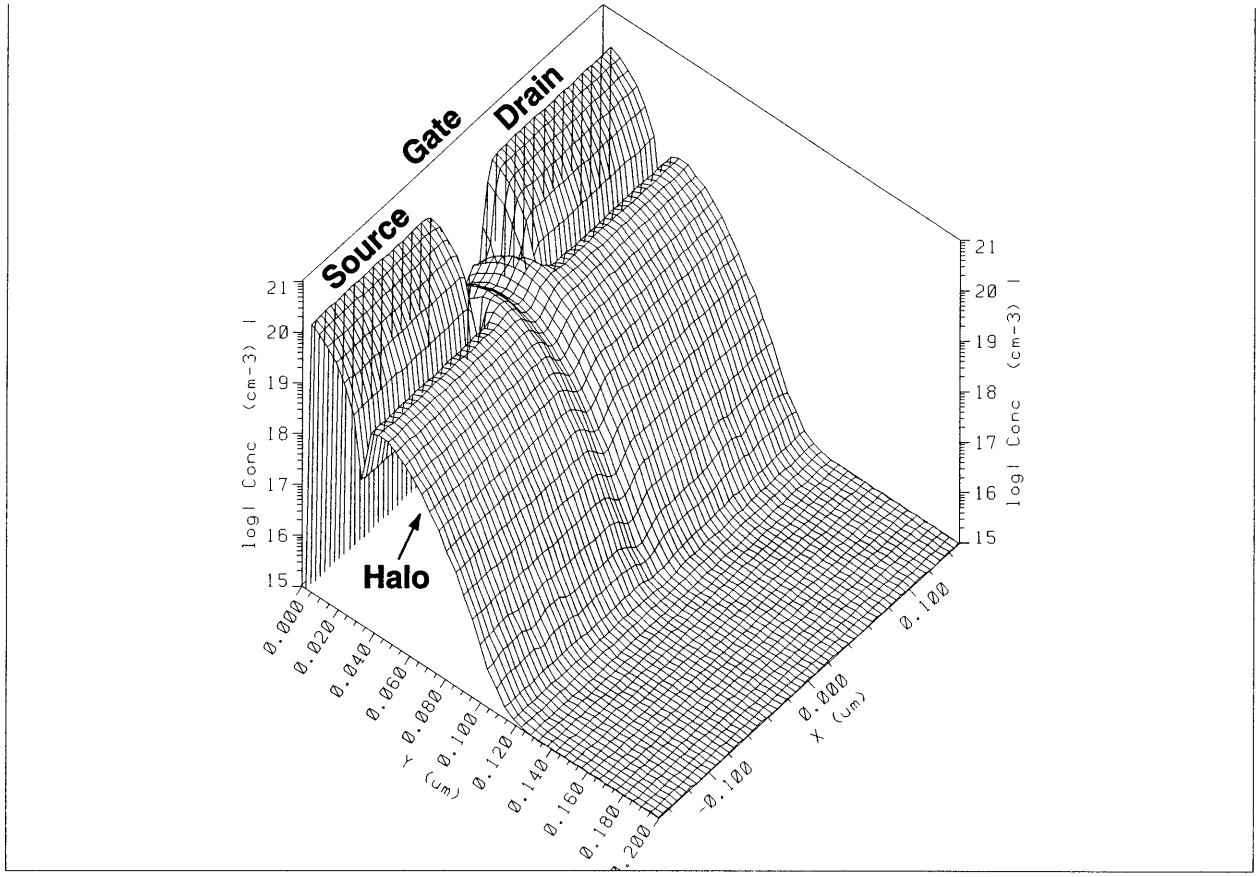


Figure 4-5. Sample net doping profile for a 100 nm technology NMOSFET with L_{GATE}=65 nm. In this example the halo was optimized at the junction depth.

physical polysilicon gate structure doped at $9x10^{19}$ cm^{-3}, and inversion layer quantization is accounted for by adding 0.3 nm to the equivalent gate oxide thickness [40].

SOI $V_{B,OFF}$, body voltage boost, and charge-balance were calculated in a similar structure, but with the addition of a deep source/drain and punchthrough doping. The deep source/drain is placed as deep as the silicon film thickness, with its lateral gaussian doping parameter set equal to that of the source/drain extension. Solutions were found using drift-diffusion modeling which simultaneously solved for potential, electron continuity, and hole continuity at each node. The models used reflect what is necessary to solve for floating-body related effects and are as follows: concentration dependent mobility for low-fields, a concentration dependent Shockley-Read-Hall model for generation and recombination with carrier lifetime set at 0.1 μs, trap-assisted band-to-band tunneling, band-gap narrowing, Auger recombination, and impact ionization. The impact

Technology	S/D s_y [nm]	$N_{A,halo}$ [cm^{-3}]	Halo s_x [nm]	Halo s_y [nm]
100 nm	19.5	$1.25x10^{19}$	17.7	25.7
70 nm	14.6	$1.36x10^{19}$	10.9	28.6
50 nm	10.1	$1.43x10^{19}$	8.3	27.2

(a)

Technology	S/D s_y [nm]	$N_{A,halo}$ [cm^{-3}]	Halo s_x [nm]	Halo s_y [nm]
100 nm	15.4	$6.14x10^{18}$	20.6	16.6
70 nm	13	$9.26x10^{18}$	12	21.9
50 nm	9.3	$1.19x10^{19}$	9.3	14.8

(b)

Table 4-2. Optimized doping profile parameters for 100, 70, and 50 nm technologies with halo set at (a) the junction depth and (b) half the junction depth.

ionization coefficients have been calibrated on the bulk MOSFET with *I-V*'s shown in Figure 2-13. The measured substrate current is matched in simulation at V_{DD}=1.01 V in a MEDICI structure where the doping profile have been inverse-modeled [41].

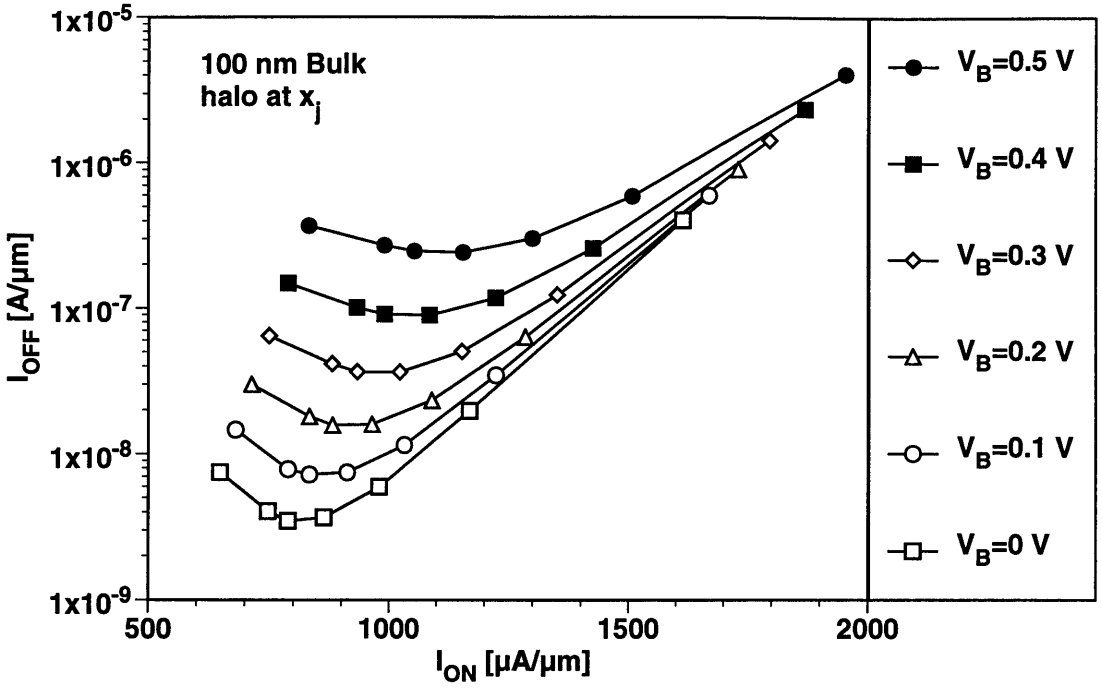
Note that solving for SOI $V_{B,OFF}$, body voltage boost, and charge-balance requires quite a different model set and solution method than solving for I_{OFF} and I_{ON}. Thus, simultaneously solving for SOI I_{OFF}, I_{ON}, $V_{B,OFF}$, and body voltage boost is much more difficult and time consuming, particularly when trying to optimize a device design. Thus, the above methodology as been developed which separates electrostatics and floating-body effects from carrier transport.

V_B [V]	at x_j [cm^{-3}]	at $x_j/2$ [cm^{-3}]
0	$1.25x10^{19}$	$6.28x10^{18}$
0.1	$1.39x10^{19}$	$6.50x10^{18}$
0.2	$1.58x10^{19}$	$7.20x10^{18}$
0.3	$1.89x10^{19}$	$8.35x10^{18}$
0.4	$2.40x10^{19}$	$1.03x10^{19}$
0.5	$3.40x10^{19}$	$1.46x10^{19}$

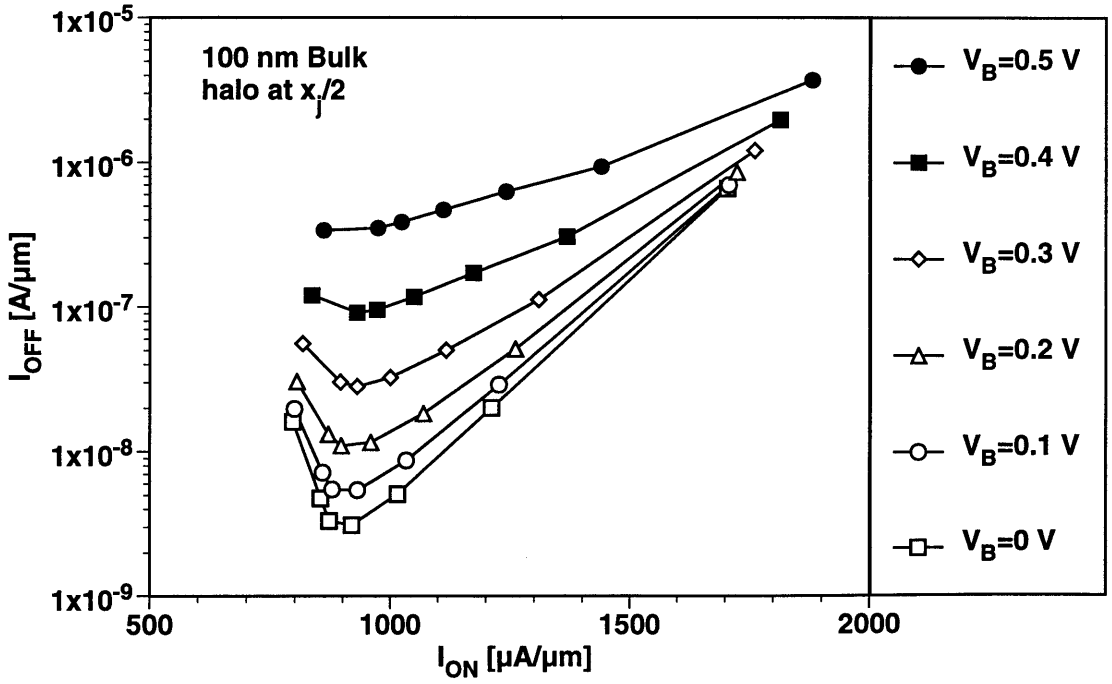
Table 4-3. Peak halo doping concentration used to match I_{OFF} at V_B to I_{OFF} at V_B=0 V for the 100 nm-technology L_{GATE}=65 nm device with the halo optimized at the junction depth and at half the junction depth.

4.4 Electrostatics

As noted above, the floating body in PD-SOI MOSFET's introduces a positive body bias in the OFF-state, $V_{B,OFF}$. Thus, the electrostatic behavior of a PD-SOI MOSFET is the same as for a bulk MOSFET with applied $V_B=V_{B,OFF}$. Figure 4-6 shows I_{OFF} versus I_{ON} for various V_B, in the range of typical $V_{B,OFF}$ values, for devices designed with 100 nm-technology parameters in Table 4-1. The V_B=0 V curve in each plot represents various L_{GATE}'s with the leftmost symbol representing L_{GATE}=100 nm, and moving right, 78, 72, 65, 58, 52, and 45 nm. The shorter L_{GATE}'s exhibit increased I_{ON} at the cost of increased I_{OFF}, with the reverse short channel effect displayed in the longer L_{GATE}'s. Applied V_B decreases threshold voltage, thereby increasing I_{OFF} and I_{ON}, shifting each point on the V_B=0 V curve up and to the right. However, the subthreshold slope is degraded due to decrease in depletion width under the gate. When the channel doping is increased (see Table 4-3) at each V_B so that $I_{OFF}(V_B)=I_{OFF}(V_B=0)$ at the 100 nm-technology nominal gate length, L_{GATE}=65 nm, the V_B>0 curves collapse into the V_B=0 curve as shown in Figure 4-7.

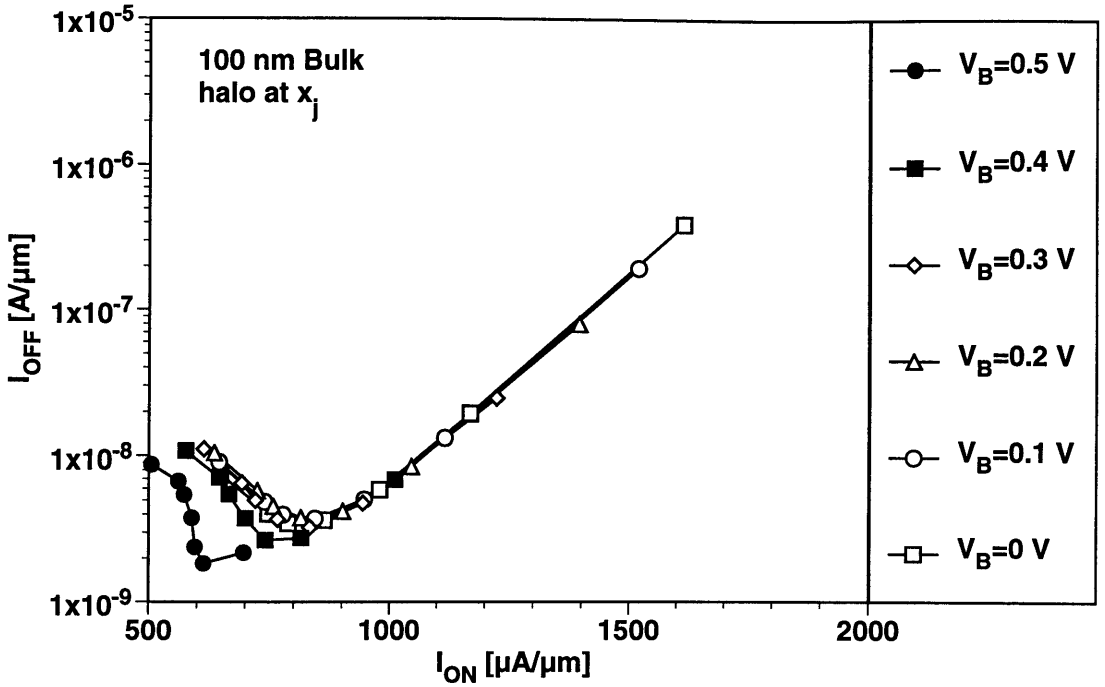


(a)

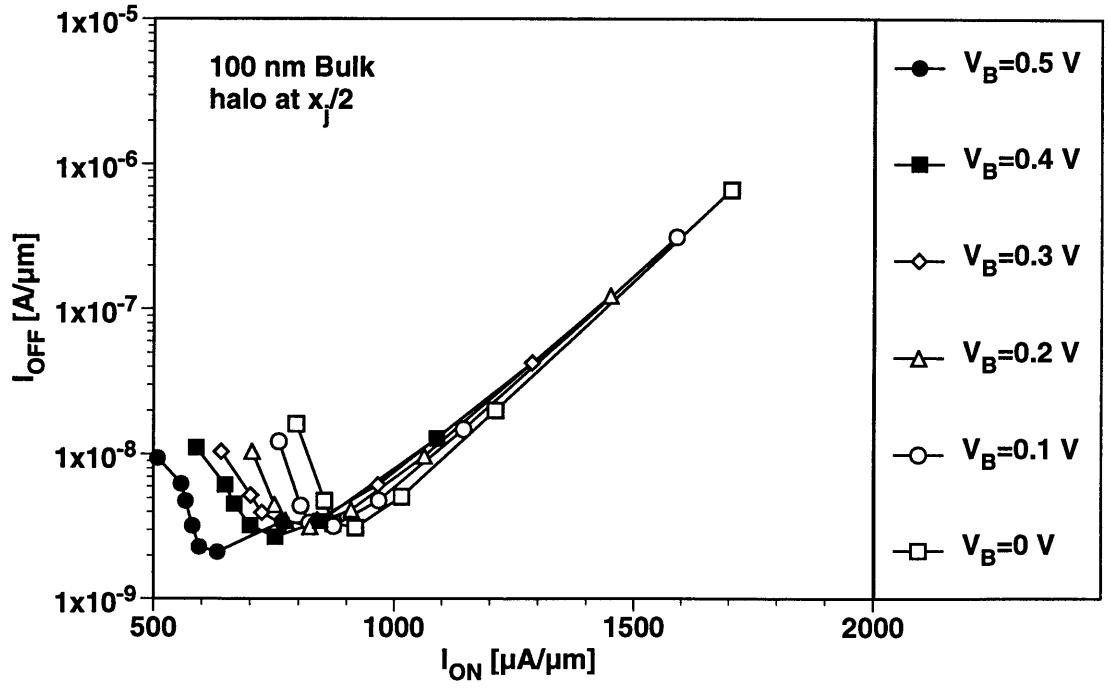


(b)

Figure 4-6. 100 nm-technology I_{OFF} versus I_{ON} for various body voltages. The peak halo doping is placed at the junction depth in (a) and at half the junction depth in (b). From left to right, the symbols on each curve correspond to L_{GATE}=100, 78,72,65,58,52, and 45 nm.



(a)



(b)

Figure 4-7. 100 nm-technology I_{OFF} versus I_{ON} for various body voltages. The peak halo doping is placed at the junction depth in (a) and at half the junction depth in (b). Doping has been modified at each body voltage so that I_{OFF} matches at the nominal gate length 65 nm. From left to right, symbols on each curve correspond to L_{GATE}=100, 78,72,65,58,52, and 45 nm.

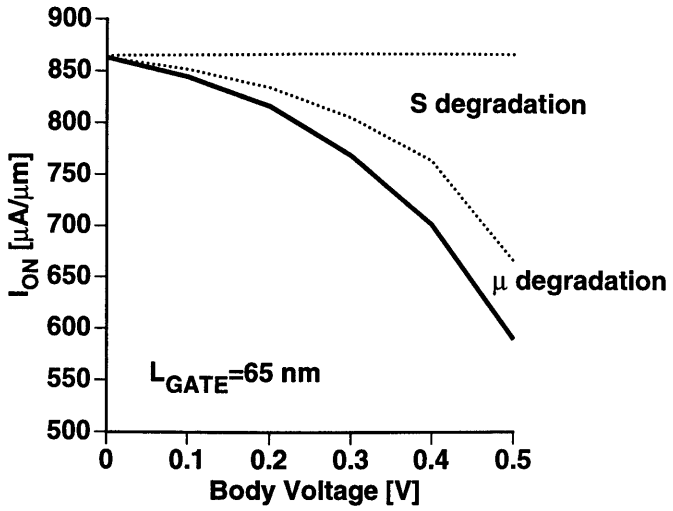


Figure 4-8. I_{ON} vs V_B for L_{GATE}=65 nm devices in figure 4-4. At each V_B, the halo doping has been increased to match I_{OFF} at V_B=0 V. I_{ON} is degraded by increased subthreshold slope and decreased mobility due to positive body voltage and increased channel doping, respectively.

Closer inspection of the plots in Figure 4-7 shows that as each of the V_B>0 curves collapse into the V_B=0 curve, the individual points shift to the left indicating a degradation in I_{ON}. Figure 4-8 shows I_{ON} degradation in the nominal gate length device in Figure 4-7 (a). Recall that I_{OFF} for V_B>0 V is matched to I_{OFF} for V_B=0 V at the nominal gate length, L_{GATE}=65 nm. Figure 4-8 shows that I_{ON} is degraded by subthreshold slope degradation due to applied V_B and mobility degradation due to the increasing channel doping to match I_{OFF}. So, from an electrostatic standpoint, positive body voltage is a disadvantage for SOI versus bulk. But for V_B≤0.3 V, the I_{ON}/I_{OFF} curve is similar to V_B=0 V and technology performance can be recovered by using shorter L_{GATE}, as positive body voltage provides better I_{OFF} control. Figure 4-9 plots I_{OFF} versus L_{GATE} showing that devices redesigned to match I_{OFF} at L_{GATE}=65 nm with applied V_B>0 V show better I_{OFF} control at shorter gate lengths. This is due to increased electrostatic integrity from positive body bias and increased channel doping.

Figure 4-10 summarizes what has been discussed by showing gate transfer *I-V* characteristics for a 100 nm-technology nominal gate length L_{GATE}=65 nm device with the doping profiles optimized with the halo at the source/drain extension junction depth. Figure 4-10 (a) shows that when V_B=0.3 V is applied, the subthreshold slope is degraded, and I_{OFF} and I_{ON} increase relative to V_B=0 V, as

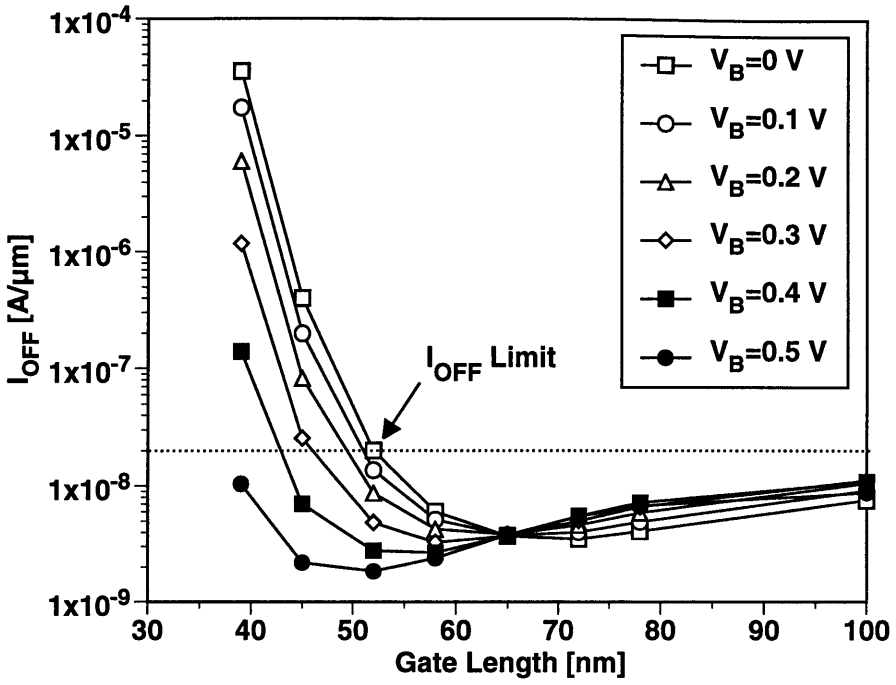
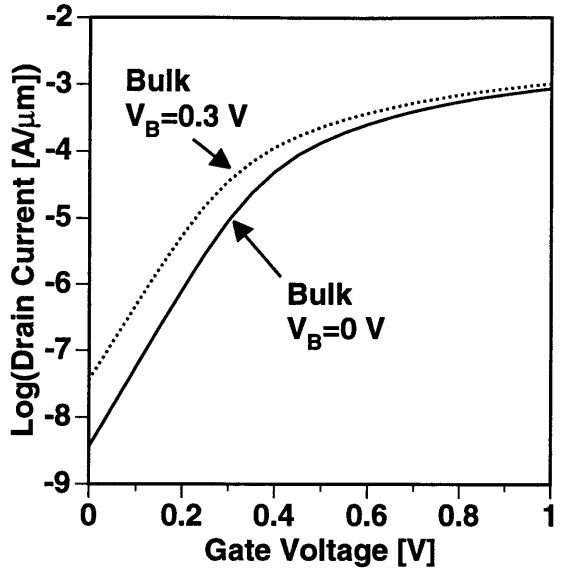


Figure 4-9. 100 nm-technology I_{OFF} vs L_{GATE} at various body voltages from figure 4-4. Short channel effects (loss of I_{OFF} control with decreasing gate length) are reduced by a combination of body voltage and increased channel doping to match I_{OFF} at the nominal gate length.

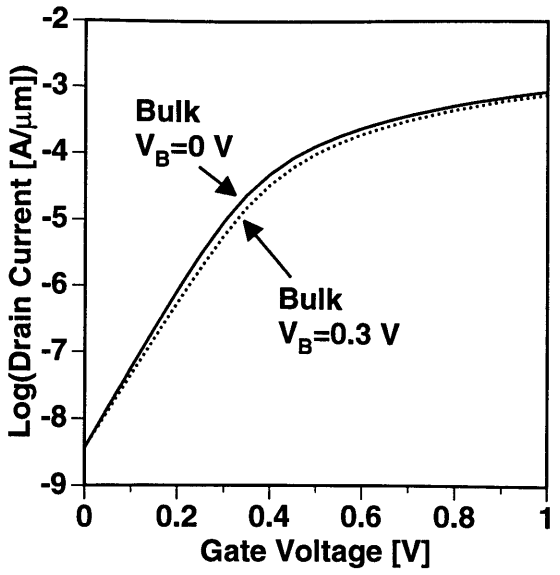
shown in Figure 4-6 (a). Figure 4-10 (b) shows that when the device at V_B=0.3 V is redesigned so that I_{OFF} at V_B=0.3 V matches the original device I_{OFF} at V_B=0 V, the *I-V* curve is degraded. I_{OFF} is has been matched, but I_{ON} is degraded, as shown in Figure 4-7 (a). When this device is simulated as an SOI structure and optimized as detailed above, $V_{B,OFF}$=0.3 V. Figure 4-10 (c) shows the *I-V* of the SOI version of this device versus the bulk. Because the body voltage is boosted by gate-body capacitive coupling in SOI, the I_{ON} is slightly better than the bulk device at V_B=0.3 V, as shown in Figure 4-10 (d). The next section details how to optimize the SOI MOSFET to maximize body voltage boost while minimizing threshold voltage hysteresis.

4.5 SOI Optimization

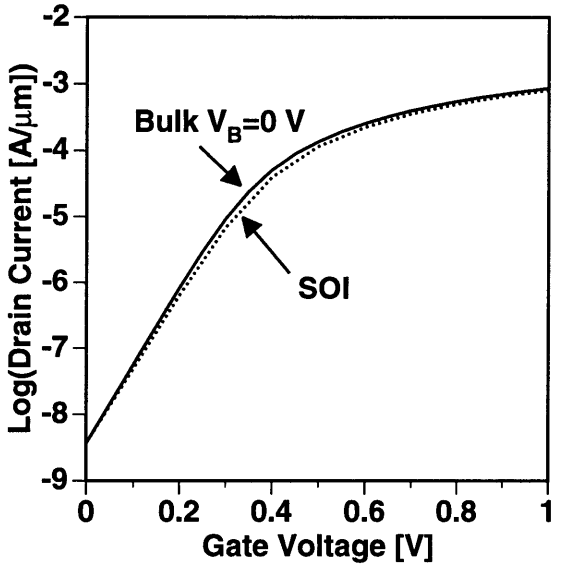
Floating-body-induced capacitive boost of I_{ON} (through body voltage boost) and threshold voltage hysteresis are strongly affected by peak halo doping depth, source/drain dimensions, and silicon film thickness. These dimensions determine the relative size of the gate-body versus body-



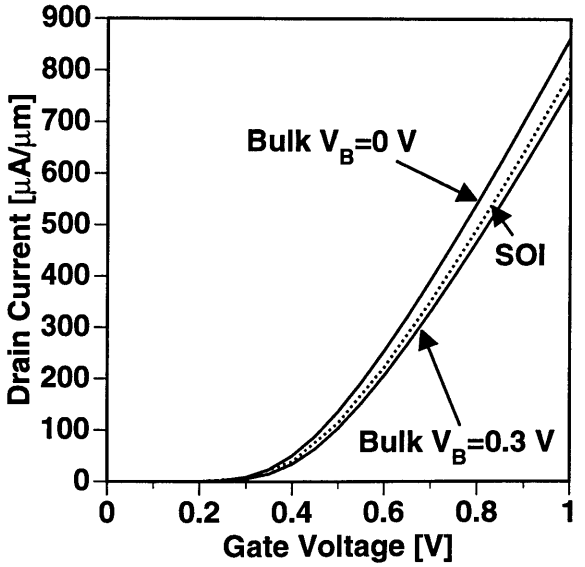
(a)



(b)



(c)



(d)

Figure 4-10. Gate transfer *I-V* characteristics for a 100 nm-technology NMOSFET with L_{GATE}=65 nm. (a) Bulk with V_B=0 V versus V_B=0.3 V at the same doping. (b) Bulk with V_B=0 V versus V_B=0.3 V, where the doping has been increased for the V_B=0.3 V case to match V_B=0 V I_{OFF}. (c) Bulk with V_B=0 V versus SOI which as $V_{B,OFF}$=0.3 V with doping adjusted in the SOI case to match bulk V_B=0 V I_{OFF}. Capacitive coupling makes SOI slightly better than bulk at V_B=0.3 V, as shown in more detail in (d).

source/drain capacitances and determine charge-balance. Within a technology, design optimization requires that these variables, shown in Figure 4-2 (b), be designed to maximize current boost and minimize hysteresis at the nominal gate length. This will minimize sensitivity to process variations.

The three variables which can be designed in the SOI structure are the source/drain extension length, the punchthrough doping level, and the silicon film thickness. They are interrelated, however. The punchthrough doping level depends on the source/drain extension length. The closer the deep source/drains are to each other, the higher the punchthrough doping required to minimize leakage from deep source to drain. The silicon film thickness for charge-balance depends on doping distribution, which is obviously affected by the source/drain extension length and punchthrough doping level. Thus, optimizing the SOI MOSFET starting from a bulk doping profile requires the following step: 1) choose a source/drain extension length, L_{ext}, and add deep source/drain, 2) optimize channel doping profile with punchthrough doping, N_{PT}, to maintain the I_{OFF}-limit at L_{MIN}, 3) design the silicon film thickness for charge-balance, $t_{si,CB}$, at L_{NOM}. Note that the punchthrough doping needs to be higher in SOI compared to bulk because of 2-D charge sharing through the buried oxide, as was shown in Figure 1-10.

Results of this procedure are shown in Figures 4-11 and 4-12, which show the result of choosing the initial halo depth at the junction depth x_j, and at half the junction depth $x_j/2$, respectively. For each L_{ext} chosen, a minimum value of N_{PT} which eliminated punchthrough leakage was found and the silicon film thickness was designed for charge-balance. Figures 4-11 and 4-12 tabulate the design parameters and show the body voltage boost versus gate length for each design. Maximized body voltage boost occurs at different L_{ext} due to the interrelationship of L_{ext}, N_{PT}, and $t_{si,CB}$ for charge-balance. This interrelationship is made more clear in Figure 4-13 which shows body voltage boost and threshold voltage hysteresis versus silicon film thickness for the nominal gate length designs with maximum body voltage boost in Figures 4-11 and 4-12. It shows that the body voltage boost is limited by the silicon film thickness for charge-balance. Recall that charge-balance at the nominal gate length minimizes threshold voltage hystersis. However, for the threshold voltage will vary between OUT-HI and OUT-LO for gate lengths which are not nominal, as shown in Figures 4-

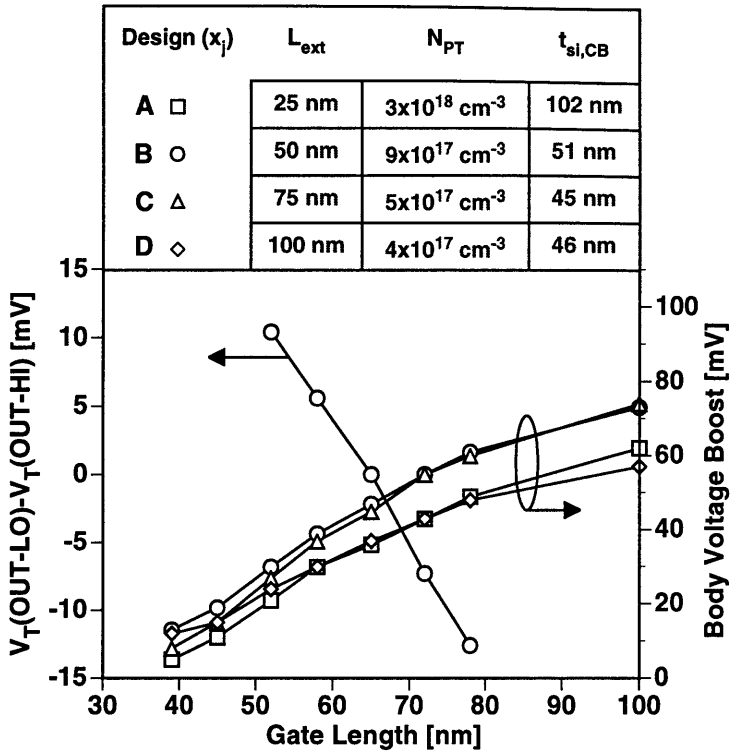


Figure 4-11. PD-SOI 100 nm-technology V_T-hysteresis and $V_{B,boost}$ vs L_{GATE} for various designs shown in above table. Peak halo doping is at the junction depth, x_j.

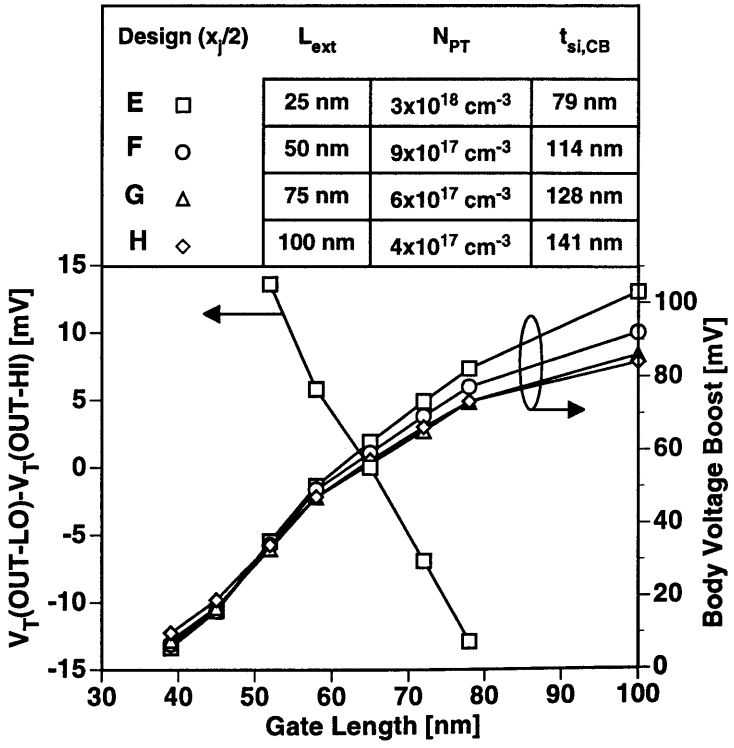


Figure 4-12. PD-SOI 100 nm-technology V_T-hysteresis and $V_{B,boost}$ vs L_{GATE} for various designs shown in above table. Peak halo doping is at half the junction depth, $x_j/2$.

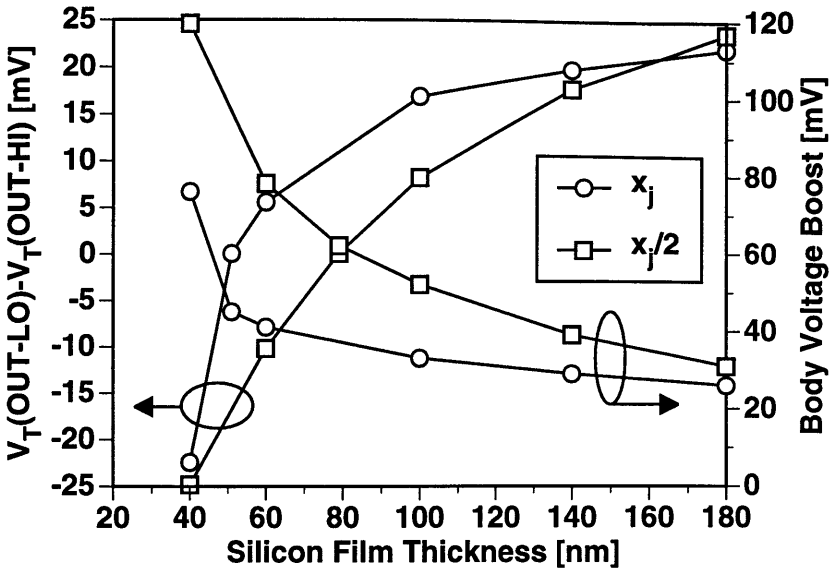


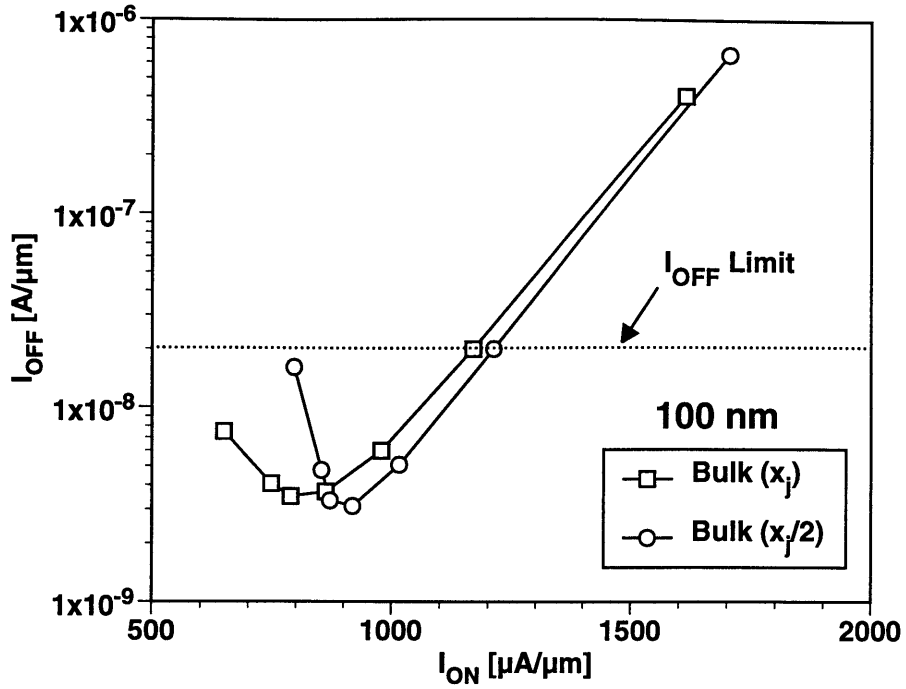
Figure 4-13. PD-SOI 100 nm-technology V_T-hysteresis and $V_{B,boost}$ vs t_{si} for peak halo doping at the junction depth (Design B in figure 4-11) and at half the junction depth (Design E in figure 4-12).

11 and 4-12.

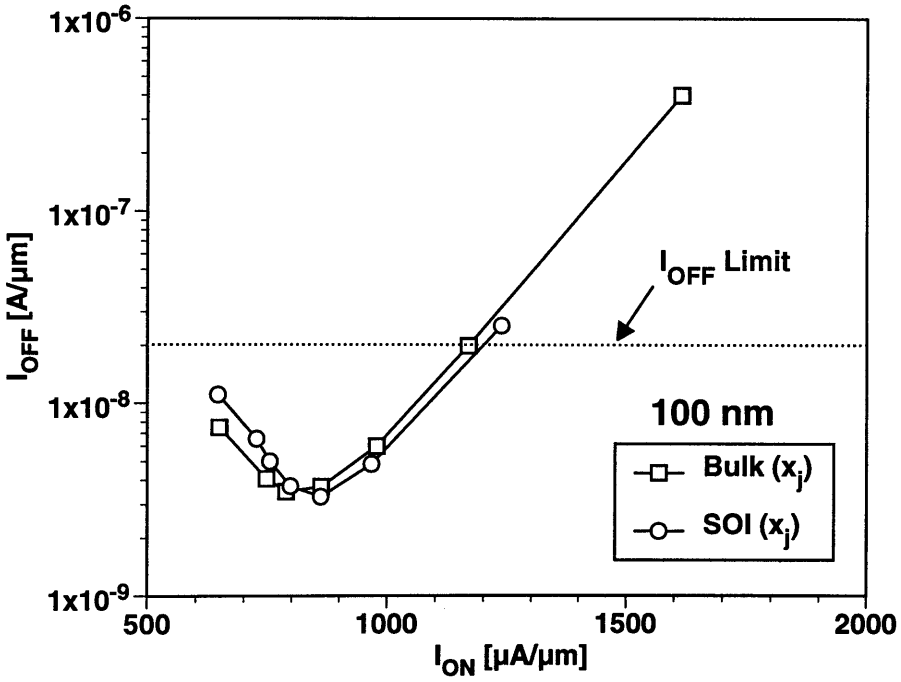
In general, halo at $x_j/2$ designs tend to have greater body voltage boost and thus greater I_{ON} than halo at x_j designs, due to decreased peak halo doping for the former. Also, as shown in Figure 4-13, $x_j/2$ designs have less sensitivity to silicon film thickness variation around the $t_{si,CB}$ compared to halo at x_j, again due to lower peak halo doping in the former.

4.6 Sub-100 nm Technologies: Bulk versus SOI

With optimized PD-SOI designs, I_{OFF} versus I_{ON} for bulk and SOI can be compared. The results are shown for 100 nm-technology in Figure 4-14. Figure 4-14 (a) shows that for bulk, placing the peak halo doping at half the source/drain extension junction depth is better than at the junction depth. This is because lower peak halo doping reduces the surface doping thereby providing higher low-field mobility and higher effective channel mobility due to lower effective perpendicular field [39]. Figure 4-14 (b) compares optimized bulk to optimized SOI in which the doping profiles have been optimized with the halo at the junction depth. In this case, SOI $V_{B,OFF}$=0.3 V. Thus the

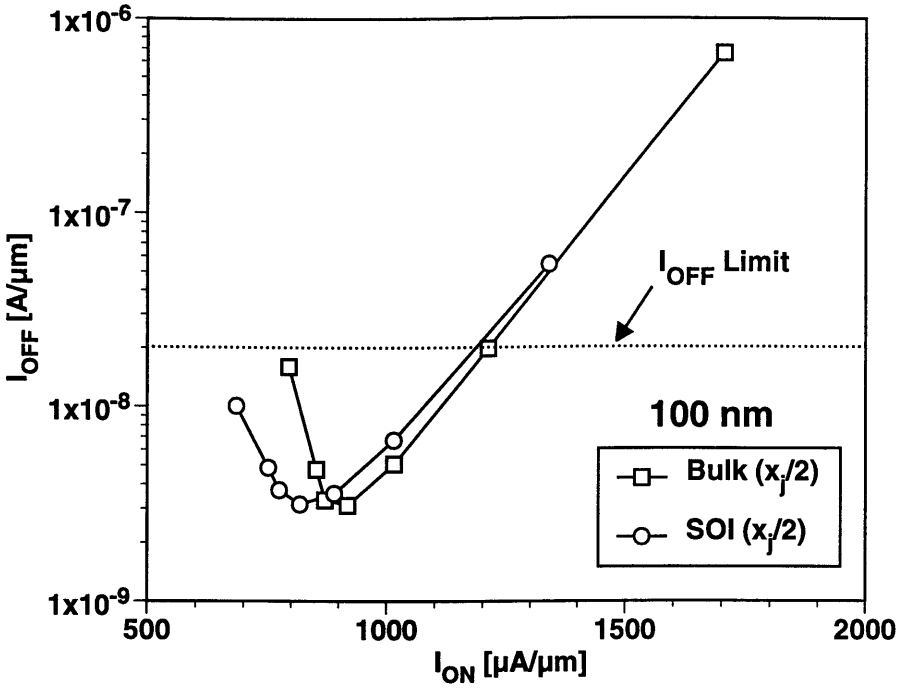


(a)

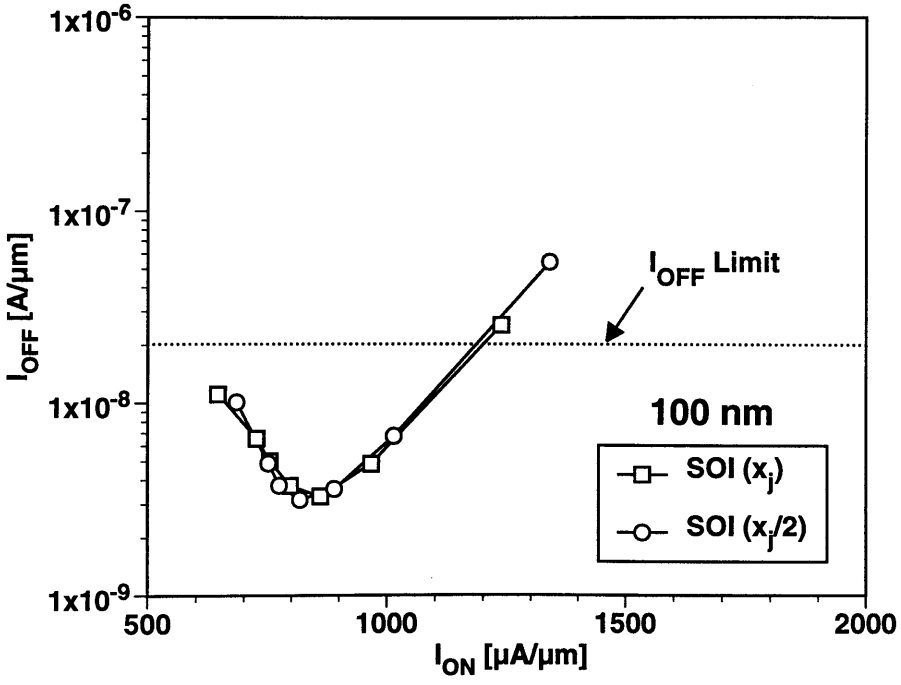


(b)

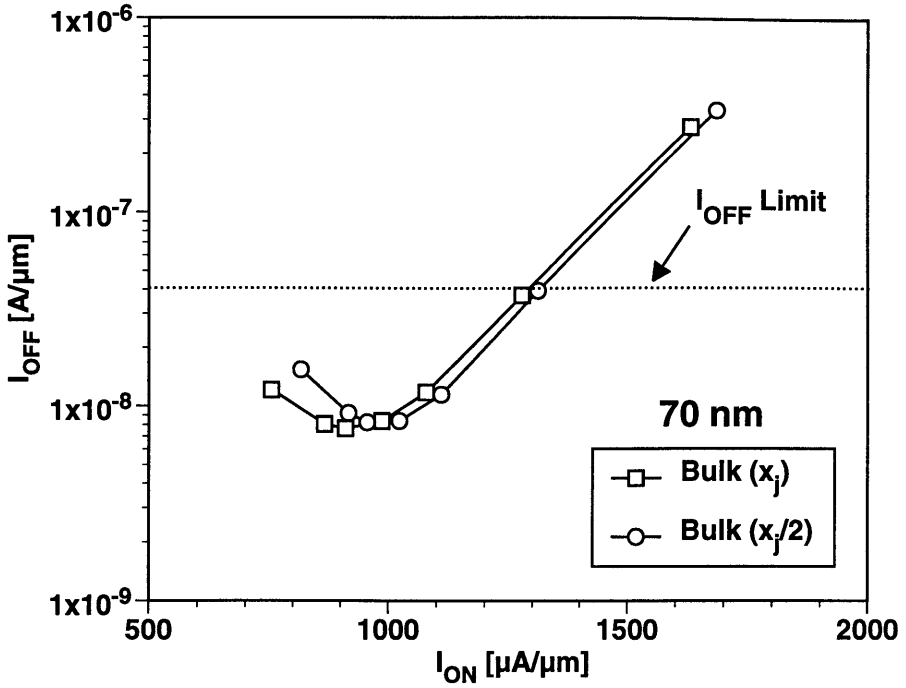
Figure 4-14. 100 nm-technology I_{OFF} versus I_{ON} for (a) bulk with halo at the junction depth versus bulk with halo at half the junction depth, (b) bulk versus SOI (Design B in figure 4-11, $V_{B,OFF}$=0.30 V) with halo at the junction depth, (c) bulk versus SOI (Design E in figure 4-12, $V_{B,OFF}$=0.27 V) with halo at half the junction depth, and (d) SOI with halo at the junction depth versus SOI with halo at the junction depth. From left to right, symbols on each curve correspond to L_{GATE}=100, 78,72,65,58,52, and 45 nm.


1x10-6
1x10-7
1x10-8
1x10-9
IOFF [A/μm]
IOFF Limit
100 nm
Bulk (xj/2)
SOI (xj/2)
500
1000
1500
2000
ION [μA/μm]

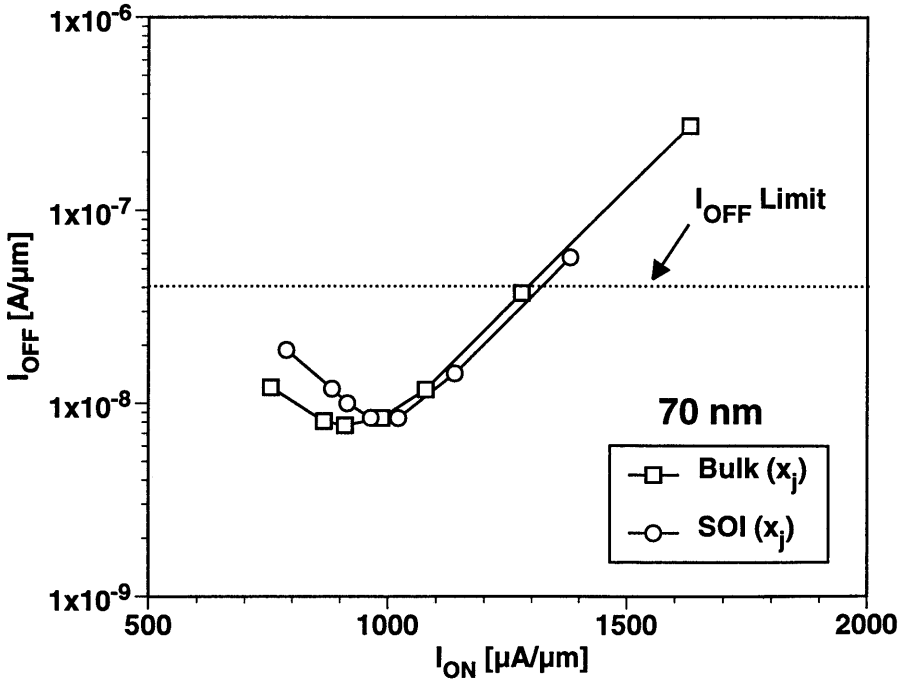
(c)


1x10-6
1x10-7
1x10-8
1x10-9
IOFF [A/μm]
IOFF Limit
100 nm
SOI (xj)
SOI (xj/2)
500
1000
1500
2000
ION [μA/μm]

(d)

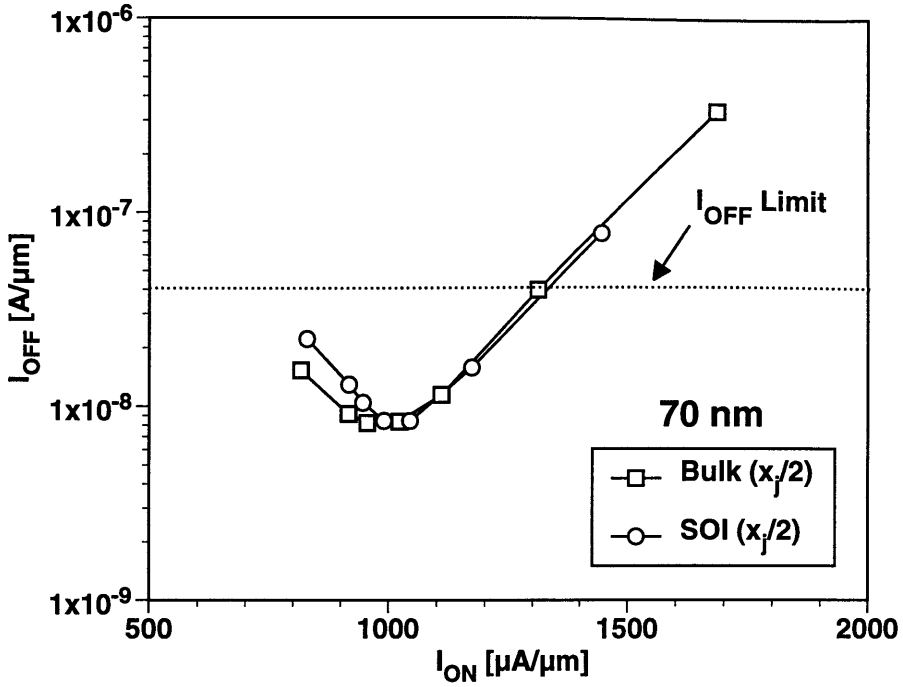


(a)

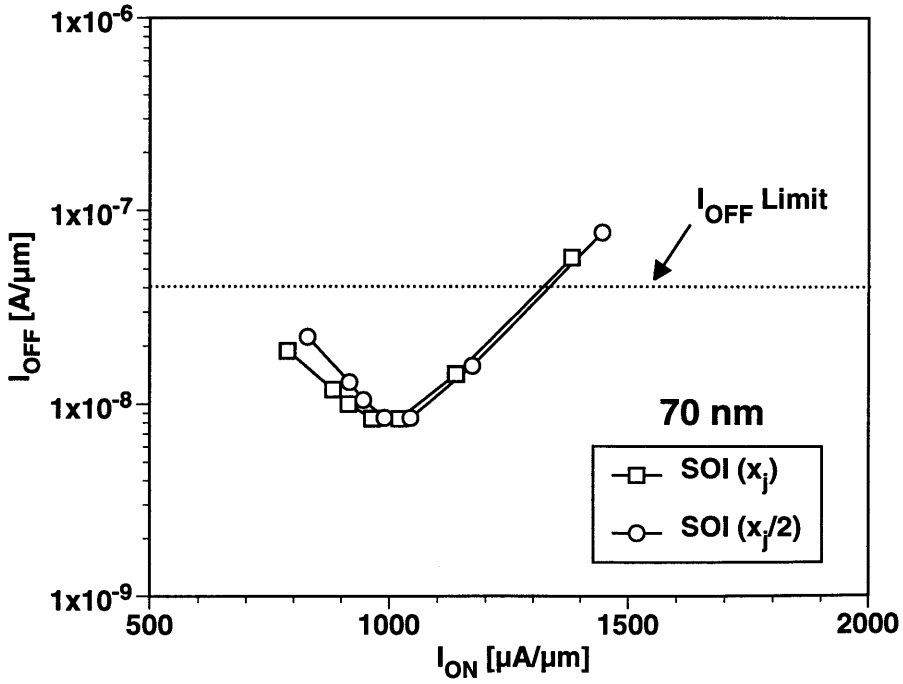


(b)

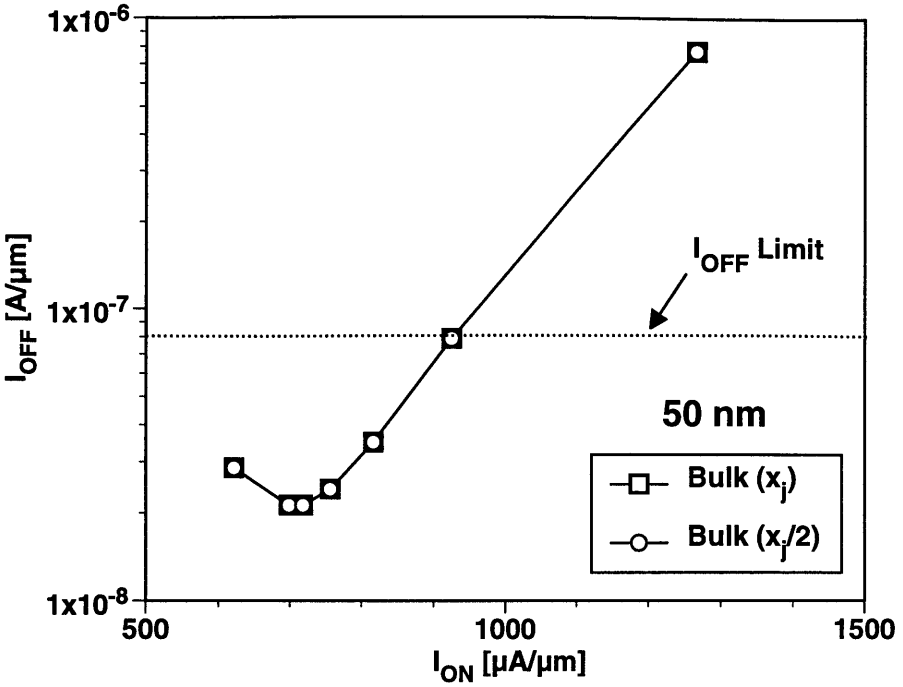
Figure 4-15. 70 nm-technology I_{OFF} versus I_{ON} for (a) bulk with halo at the junction depth versus bulk with halo at half the junction depth, (b) bulk versus SOI (L_{ext}=35 nm, N_{PT}=3x10^{18} cm^{-3}, $t_{si,CB}$=38 nm, $V_{B,OFF}$=0.28 V) with halo at the junction depth, (c) bulk versus SOI (L_{ext}=25 nm, N_{PT}=3x10^{18} cm^{-3}, $t_{si,CB}$=70 nm, $V_{B,OFF}$=0.26 V) with halo at half the junction depth, and (d) SOI with halo at the junction depth versus SOI with halo at the junction depth. From left to right, symbols on each curve correspond to L_{GATE}=70, 54,50,45,41,36, and 27 nm.


1x10-6
1x10-7
1x10-8
1x10-9
IOFF [A/μm]
IOFF Limit
70 nm
Bulk (xj/2)
SOI (xj/2)
500
1000
1500
2000
ION [μA/μm]

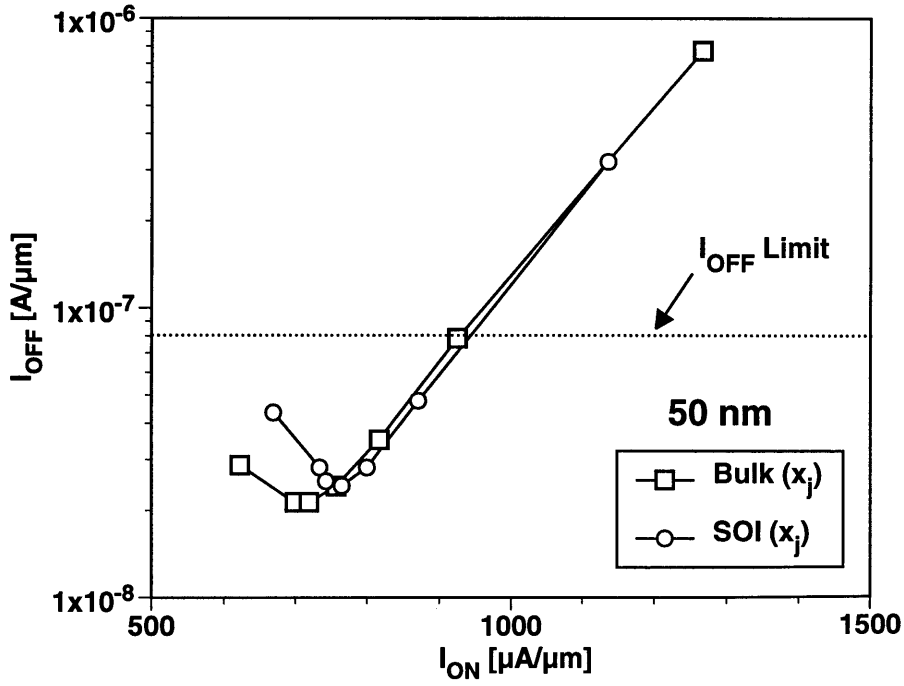
(c)


1x10-6
1x10-7
1x10-8
1x10-9
IOFF [A/μm]
IOFF Limit
70 nm
SOI (xj)
SOI (xj/2)
500
1000
1500
2000
ION [μA/μm]

(d)

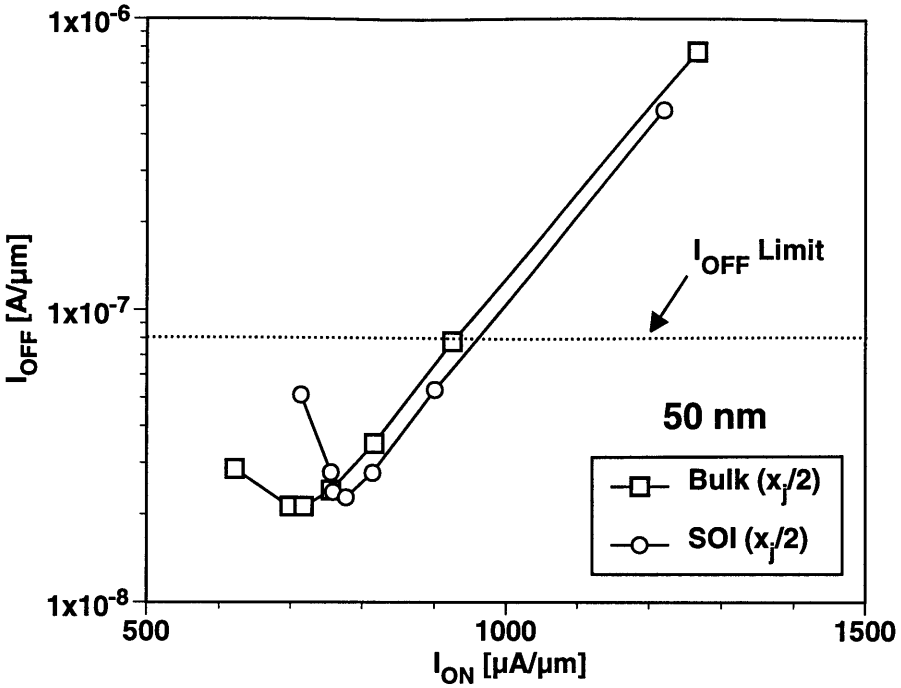


(a)

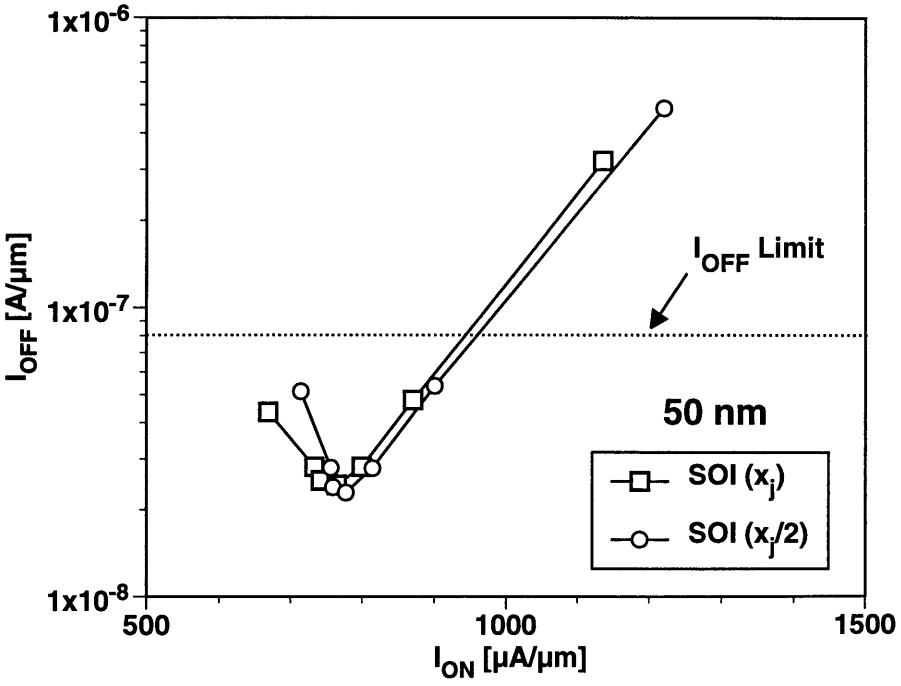


(b)

Figure 4-16. 50 nm-technology I_{OFF} versus I_{ON} for (a) bulk with halo at the junction depth versus bulk with halo at half the junction depth, (b) bulk versus SOI (L_{ext}=25 nm, N_{PT}=4x10^{18} cm^{-3}, $t_{si,CB}$=30 nm, $V_{B,OFF}$=0.23 V) with halo at the junction depth, (c) bulk versus SOI (L_{ext}=20 nm, N_{PT}=5x10^{18} cm^{-3}, $t_{si,CB}$=65 nm, $V_{B,OFF}$=0.21 V) with halo at half the junction depth, and (d) SOI with halo at the junction depth versus SOI with halo at the junction depth. From left to right, symbols on each curve correspond to L_{GATE}=50, 38,35,32,29,26, and 22 nm.


1x10-6
1x10-7
1x10-8
IOFF [A/μm]
IOFF Limit
50 nm
Bulk (xj/2)
SOI (xj/2)
500
1000
1500
ION [μA/μm]

(c)


1x10-6
1x10-7
1x10-8
IOFF [A/μm]
IOFF Limit
50 nm
SOI (xj)
SOI (xj/2)
500
1000
1500
ION [μA/μm]

(d)

SOI I_{ON} is degraded compared to bulk which the body voltage boost cannot recover. The added electrostatic integrity for SOI with $V_{B,OFF}$=0.3 V, however, means that the nominal gate length can be made shorter, because shortest SOI devices shown is still near the I_{OFF}-limit. In this way, SOI can recover I_{ON}. In addition, using shorter gate lengths reduces capacitive load. Thus, using the common metric CV_{DD}/I_{ON} [42], SOI can achieve similar performance to bulk by using shorter gate lengths.

Figure 4-14 (c) shows a similar result, comparing bulk and SOI optimized with the halo at half the junction depth. Figure 4-14 (d) compares SOI with the halo placed at the junction depth and at half the junction depth. The two curves are similar, although the SOI with halo at half the junction depth has more I_{ON}, due to greater body voltage boost, as shown in Figure 4-13. The same optimization is applied to the 70 nm- and 50 nm-technology nodes and the results are shown in Figures 4-15 and 4-16. These results show that PD-SOI I_{ON}/I_{OFF} is similar to bulk and that shallower halo placement increases I_{ON} current boost and requires thicker and therefore more manufacturable silicon film thicknesses. In general, as the power supply is reduce, $V_{B,OFF}$ is reduced which makes SOI yet more similar to bulk.

4.7 Summary

This chapter presented a PD-SOI design and optimization framework which showed that with proper design, PD-SOI I_{ON}/I_{OFF} is comparable to bulk in sub-100 nm CMOS technologies. I_{ON} at each L_{GATE} is degraded due to nonzero body bias but performance metrics for the technology as a whole (such as CV_{DD}/I_{ON}) can be recovered by using shorter L_{GATE} as L_{NOM}. In this way, intrinsic SOI MOSFET performance can be similar to bulk. In addition, PD-SOI still has the advantage over bulk with reduced parasitic source/drain capacitance and circuit-level advantages. While it has been noted that the reduced-parasitics advantage decreases with scaling [43], circuit-level advantages such as reduction of and reverse body effect in stacked transistors will increase as power supplies are scaled and gate overdrive is reduced [38].

Chapter 5

Double-Gate MOSFET Technology

As noted in Chapter 1, double-gate fully-depleted SOI MOSFET's are expected to scale to shorter gate lengths than both partially-depleted SOI or bulk-silicon MOSFET's. However, fabrication of a bottom-gate which is well-aligned to the top-gate is challenging. The test vehicle for investigating double-gate fabrication technology is an optimized version of SOIAS (silicon-on-insulator with active substrate) CMOS technology. This technology was developed by Yang, et al., and uses the bottom-gate to shift the threshold voltage low during periods of operation, and high during idle periods [44].

The bottom-gate in SOIAS technology was patterned into a buried undoped polysilicon layer through the active region. This resulted in a large parasitic capacitance from source/drain to the bottom-gate. The optimized SOIAS technology studied in this chapter, aligns MOSFET's to buried patterned bottom-gates, as illustrated in Figure 5-1. This significantly reduces the source/drain-bottom-gate capacitance. The bottom-gates are oversized with respect to the top-gates by alignment tolerances. A counterdoping implant can be done through the MOSFET top-gate to minimize the parasitic capacitance from the source/drain to the oversized bottom-gate, thereby self-aligning the bottom-gate to the top-gate.

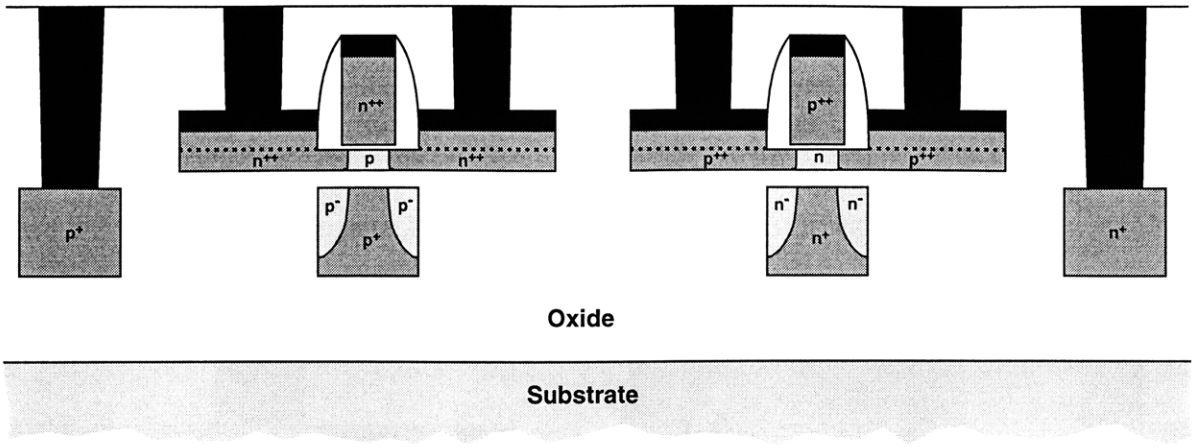


Figure 5-1. Schematic cross section of proposed self-aligned double-gate CMOS. Black represents metal and salicide and the dashed lines denote epitaxially grown raised source/drain.

5.1 Process Flow

Process development focused on the NMOSFET. The process flow to fabricate NMOSFET's discussed in section 5.3 is shown in Figure 5-2 (a)-(m). The starting material is a SOI wafer. SIMOX wafers with silicon film thickness of 200 nm and buried oxide thickness of 380 nm were used. After thinning the silicon film by dry oxidation and HF-dip, the bottom-gate is formed as shown in Figure 5-2 (b). The bottom-gate material is deposited as amorphous silicon at 560 °C, implanted with boron, patterned and etched, covered with LTO, and then recrystallized at 600 °C for 14 hours into large-grain polysilicon [45]. The LTO is then densified at 1000 °C for 10 minutes before chemical-mechanical-polish (CMP), which will planarize the LTO as shown in Figure 5-2 (c).

The SOI wafer is then flipped and bonded to a bulk-silicon handle wafer as shown in Figure 5-2 (d). The handle wafer has 500 nm thermal oxide grown on it, so this is an oxide-oxide bond. The bond is annealed at 950 °C for 40 minutes. Note that the bonding temperature and all subsequent temperatures in the process should be below the LTO densification temperature to insure no further outgassing from the LTO, which may affect bond integrity. The substrate of the SOI wafer is then

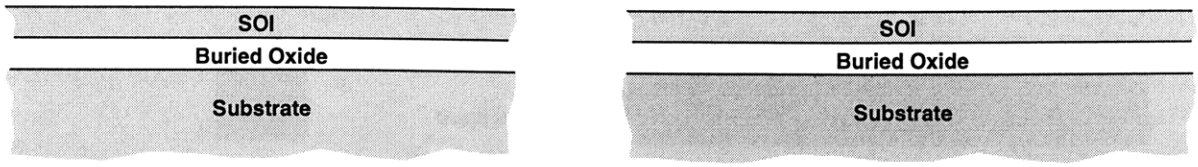


Figure 5-2 (a) Cross section and side view of starting starting SOI wafer.

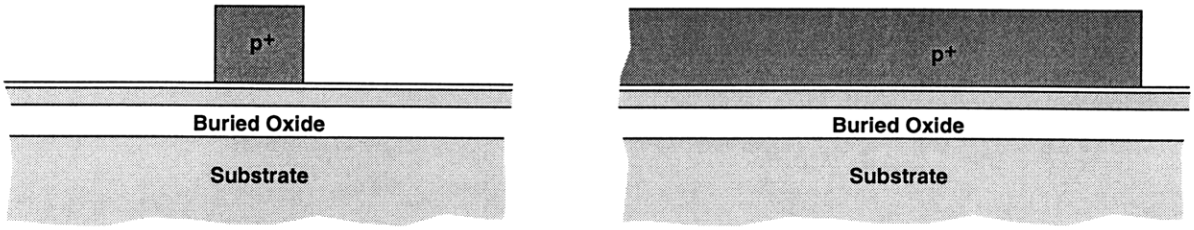


Figure 5-2 (b) Cross section and side view. The bottom-gate oxide is grown on the SOI. The bottom-gate is formed by deposition of polysilicon, boron implant, patterning, and etching.

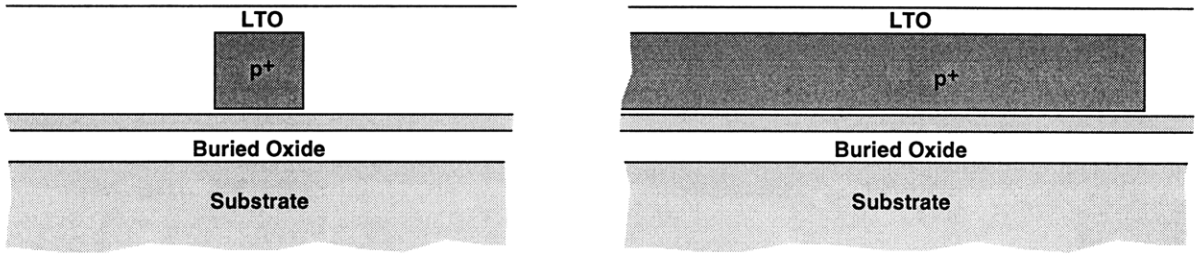


Figure 5-2 (c) Cross section and side view. LTO is deposited over the bottom-gate and is densified. The LTO is then chemically-mechanically-polished flat.

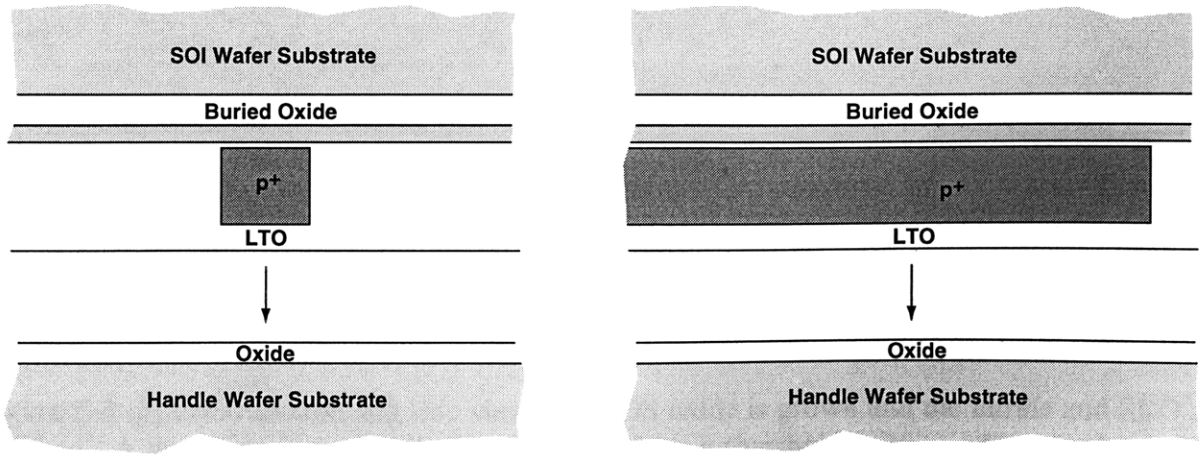


Figure 5-2 (d) Cross section and side view. The SOI wafer is flipped and bonded to a handle wafer.

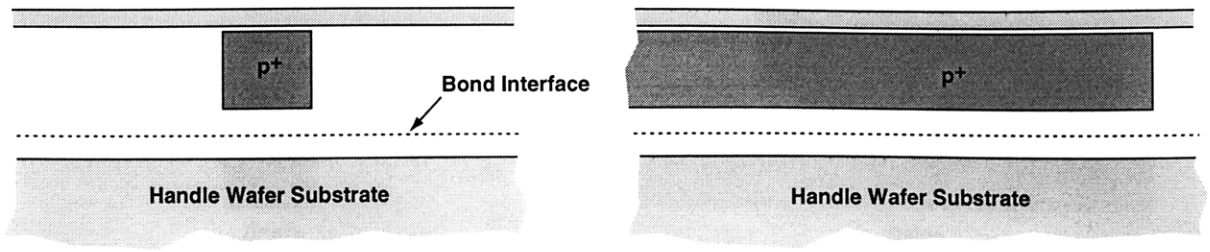


Figure 5-2 (e) Cross section and side view. The SOI wafers substrate and buried oxide are removed.

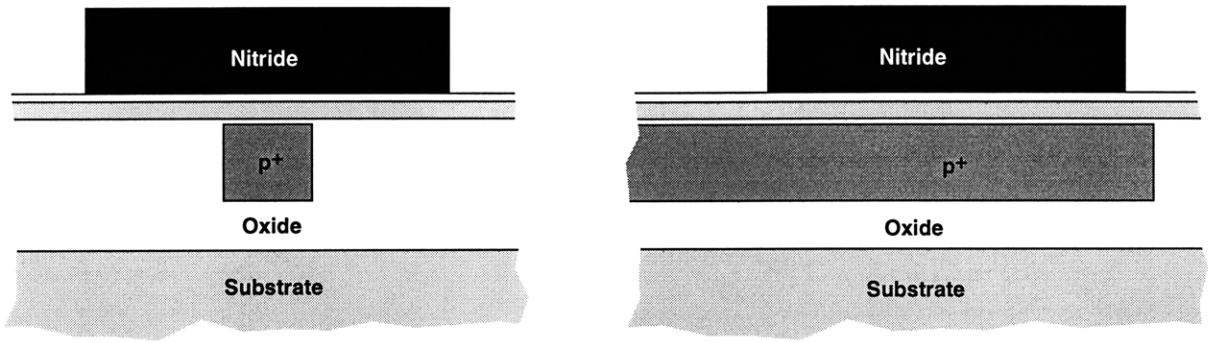


Figure 5-2 (f) Cross section and side view. LOCOS isolation definition. A SRO is grown and nitride is deposited, patterned, and etched. A field implant is performed.

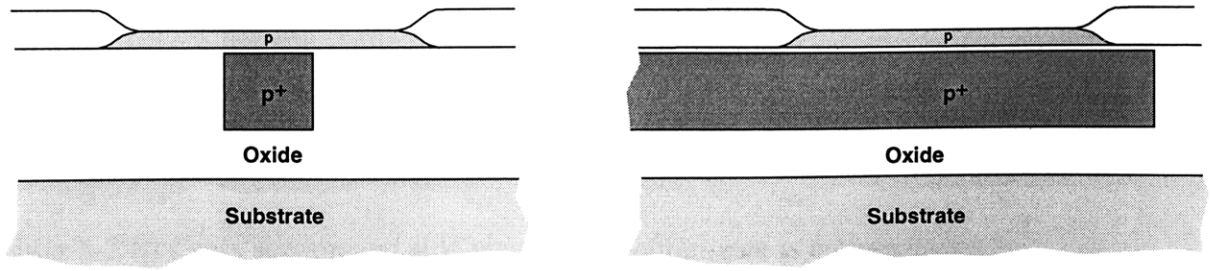


Figure 5-2 (g) Cross section and side view. LOCOS oxide is grown and the nitride and SRO are stripped. A dummy gate oxide is grown followed by a threshold voltage adjust implant. The dummy gate oxide is the stripped.

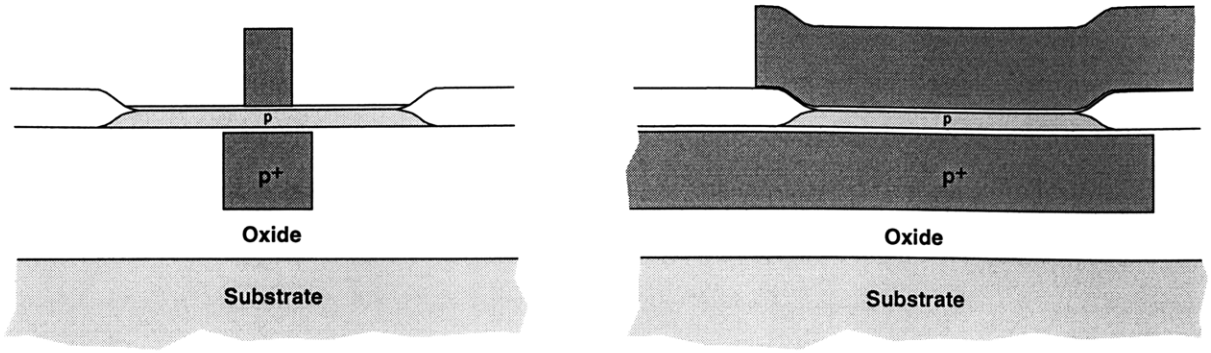


Figure 5-2 (h) Cross section and side view. The gate oxide is grown and polysilicon is deposited. The polysilicon is patterned and etched to form the top-gate.

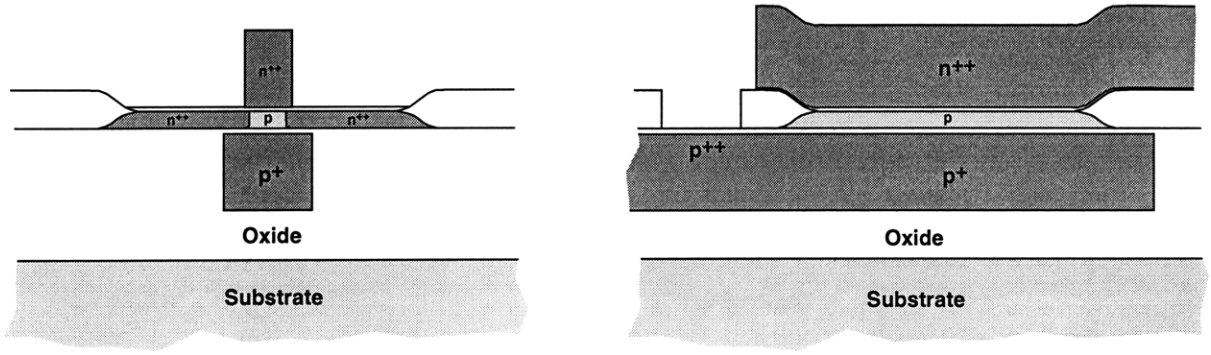


Figure 5-2 (i) Cross section and side view. The top-gate and source drain are implanted. A contact cut to the bottom-gate is patterned and etched. The bottom-gate contact cut is implanted to assure ohmic behavior.

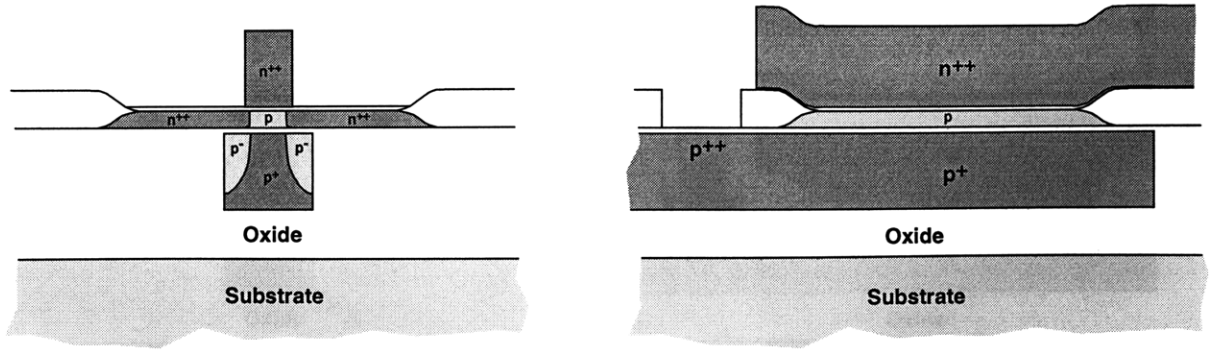


Figure 5-2 (j) Cross section and side view. The bottom-gate is counterdoped to reduce parasitic source/drain-bottom-gate capacitance. Note that a resist mask protects the bottom-gate contact regions during the counterdoping.

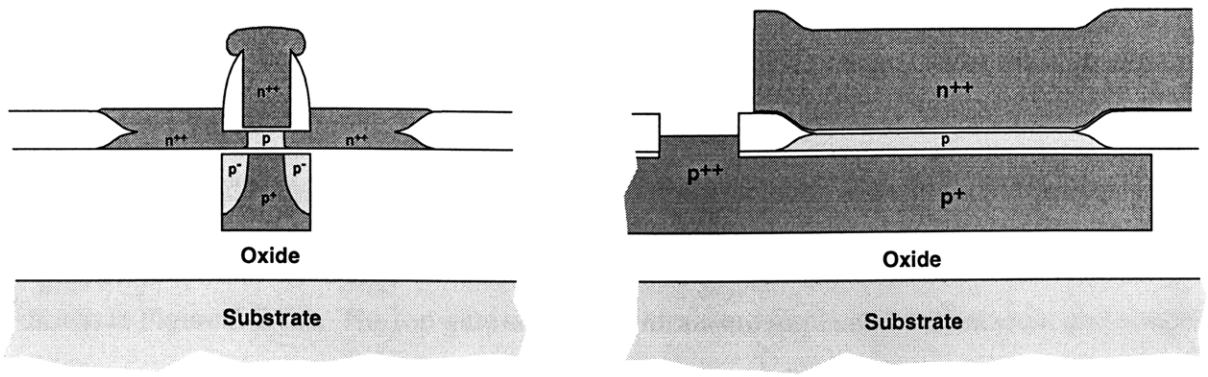


Figure 5-2 (k) Cross section and side view. A spacer is formed. A raised source/drain is grown by UHVCVD selective epitaxial growth.

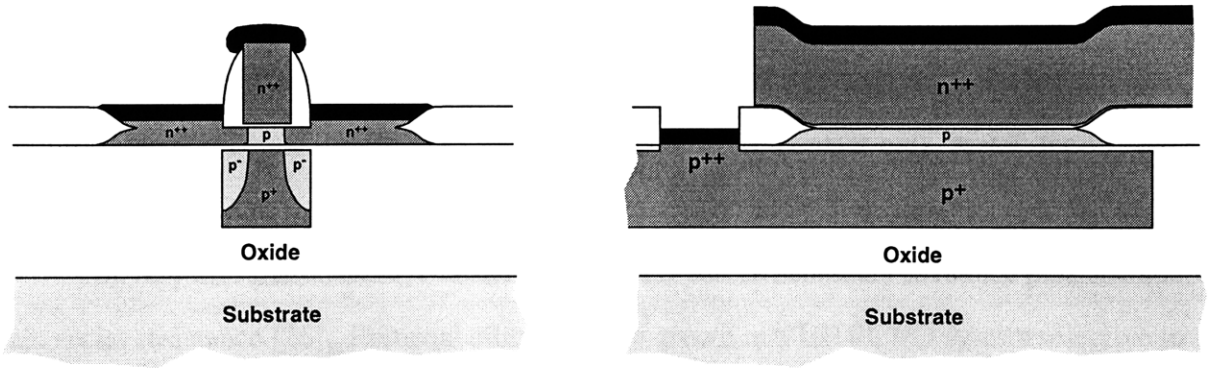


Figure 5-2 (l). Cross section and side view. The raised source/drain is salicided.

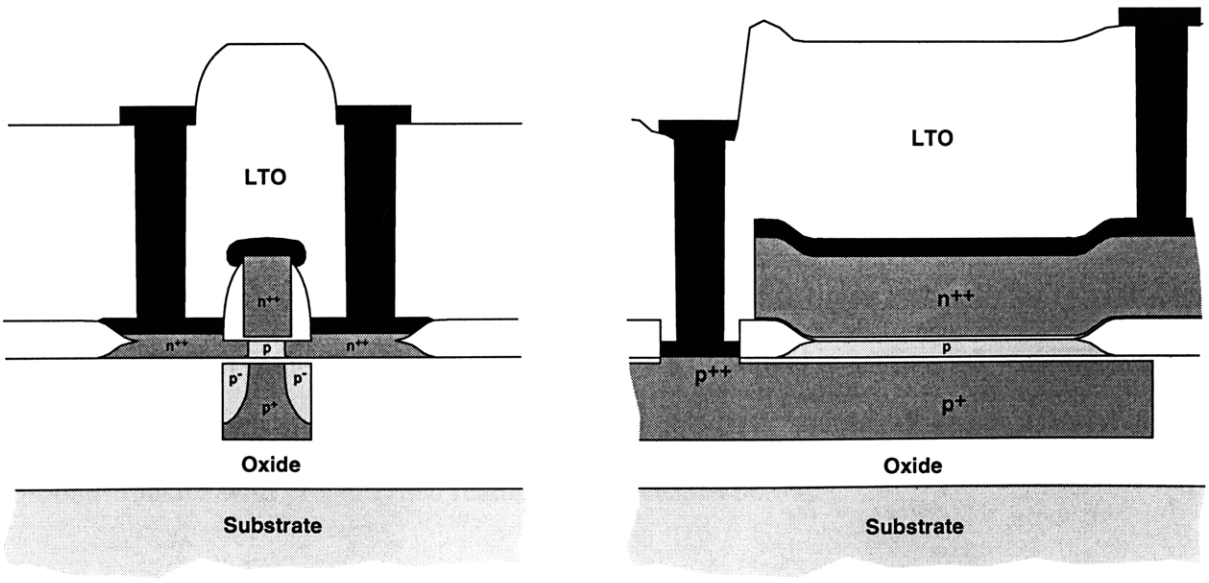


Figure 5-2 (m) Cross section and side view. LTO is deposited and contact cuts are patterned and etched. Metal is deposited, patterend, and etched.

removed by mechanical grinding and chemical silicon etch, with the buried oxide layer serving as an etchstop. Once the buried oxide layer is removed, a NMOSFET process can be run with alignment to the bottom-gate, now buried, as shown in Figure 5-2 (e).

The standard NMOSFET process includes LOCOS isolation to define active areas, as shown in Figures 5-2 (f) and (g). The top-gate stack is formed and patterned, aligned to the bottom-gate, as shown in Figure 5-2 (h). The top-gate and source/drain are implanted with arsenic and annealed at 900 °C for 20 minutes, as shown in Figure 5-2 (i). This is followed by an arsenic implant at higher energy which will counterdope the bottom-gate regions which overlap the source/drain as shown in Figure 5-2 (j). The counterdope implant is annealed at 600 °C to minimize lateral diffusion of the implant. Also shown in Figure 5-2 (j) is the formation of a contact cut to the bottom-gate. At this point, the device can be finished with LTO passivation, contact cuts, metal deposition, and metal patterning.

For very thin silicon films, a raised source/drain will be necessary to reduce parasitic source/drain series resistance [26]. Epitaxial silicon can be grown in a UHVCVD system selective to the spacer and the field oxide to form the raised source/drain, as shown in Figure 5-2 (k). Once the raised source/drain is grown, it can be silicided as shown in Figure 5-2 (l). The device can then be finished with LTO passivation, contact cuts, metal deposition, and metal patterning.

For full details of each process step, please refer to Appendix B. The raised source/drain process was not completed for this work because the silicon epitaxial growth available was not selective to silicon nitride spacer material. Oxide spacers can be used, but are difficult to fabricate and are not compatible with thin-film silicon processing because oxide etch selectivity to silicon is difficult to establish [26]. The double-gate MOSFET's presented in section 5.3 were fabricated by completing the process after the step illustrated in Figure 5-2 (j). This result demonstrates successful flip-bond-transfer with prepatterned features, as discussed below.

5.2 Flip-Bond-Transfer Process

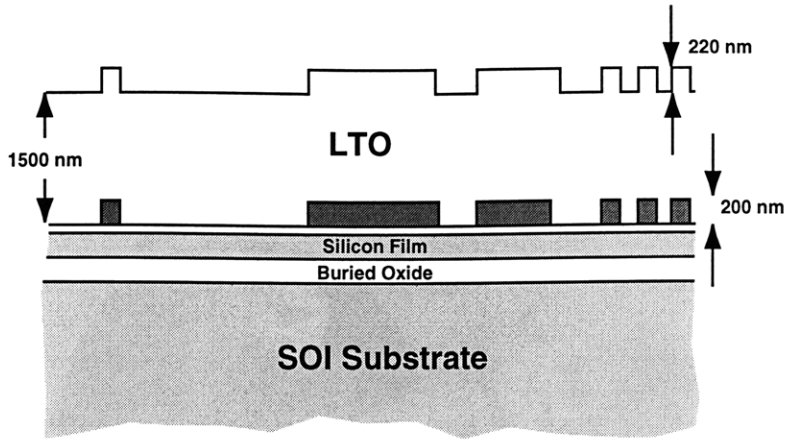
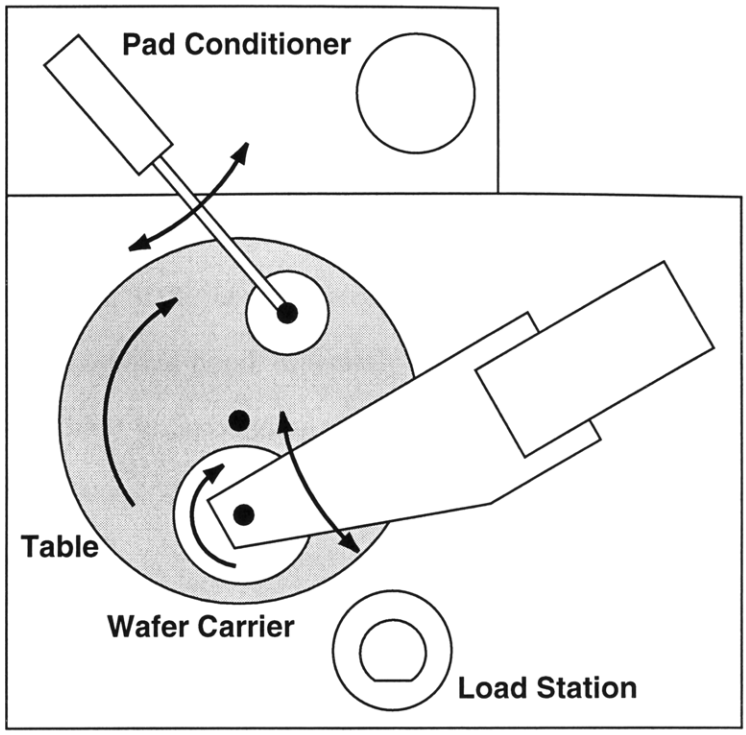


Figure 5-3. Cross section of bottom-gate stack prior to CMP. The CMP process was optimized to flatten a polysilicon step-height of 200 nm. Deposition of 1.5 μm LTO resulted in a step-height of 220 nm.

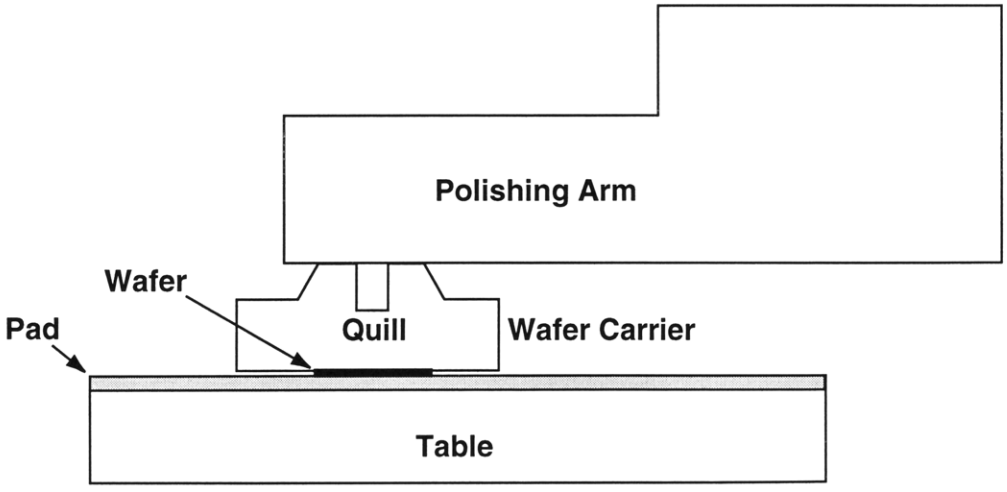
The flip-bond-transfer process presented here is used to transfer patterned features from a SOI wafer to another wafer. The patterned stack for which this process was optimized is shown in Figure 5-3. It consists of patterned 200 nm thick polysilicon bottom-gates on bottom-gate oxide on SOI. These are covered by 1.5 μm LTO, which has a resulting feature step height of 220 nm. Flip-bond-transfer has three steps, the first of which is to flatten the features on the LTO by chemical-mechanical-polish. The second step is to bond the SOI wafer to a handle wafer, and the third step is to remove the SOI wafer substrate.

5.2.1 Chemical-Mechanical-Polishing (CMP)

Chemical-mechanical-polishing (CMP) uses mechanical motion of a polishing pad in combination with a slurry of chemicals and particles to polish surfaces. This technology has been extensively used throughout history to polish glass. CMP was introduced to ULSI manufacturing as a method to planarize dielectrics between levels of metal. This allowed more metal layers to be integrated into multilevel metal processes. Since then, CMP has found its way into polishing dielectrics for isolation between devices (shallow trench isolation), and into polishing of metal to



(a)



(b)

Figure 5-4. (a) Schematic top view of the Strausbaugh 6EC CMP machine. Black dots indicate axes of rotation. Arrows indicate direction of rotation. Double-headed arrows indicate directions of movement. (b) Side view of the wafer carrier, wafer, table, and pad. The wafer carrier rotates round the quill.

define metal lines in multilevel interconnect (damascene metal process) [46].

The flip-bond-transfer process requires the CMP system to polish LTO features flat for wafer bonding. The wafer must be both locally (feature-level) and globally (die- and wafer-level) flat. If it is not locally flat, the wafers will not bond. If it is not globally flat, air pockets can form during bonding. In addition, wafer bonding requires that the surface be as smooth and clean as possible for the strongest possible bond, especially at lower bonding temperatures, <1100 °C for oxide-oxide bond [47]. These wafer bonding requirements are more stringent than those of typical CMP applications listed above. Thus, very careful attention must be paid when planarizing and/or polishing for wafer bonding.

The machine used is in this work a Strasbaugh Model 6EC Laboratory Planarizer [48]. The pad used is a Rodel P05677 and the slurry used is Cabot Semi-Sperse 25. This pad and slurry combination is designed for polishing of oxide surfaces [48]. The pad is of polyethylene, softer than the oxide being polished. The slurry is made of 100 nm sized silica particles suspended in a KOH-based solution. The pad by itself cannot polish or planarize the oxide without the slurry, hence the chemical-mechanical process.

Figure 4-5 shows a top-view of the machine and a side-view of the polishing system. The wafer is placed with the side to be polished facedown into the load station. The polish arm picks up the wafer into the wafer carrier and moves it to the polishing table, but is tilted slightly to keep the wafer suspended above the table. At this point the table begins to rotate and the wafer carrier begins to rotate around the quill. The direction of rotation is shown in Figure 5-4 (a). Slurry also starts being dispensed onto the polishing table as soon as the table and wafer carrier begin rotating. The polishing arm then pushes the wafer into contact with the polishing pad. While polishing, the polishing arm moves from side to side as indicated in Figure 5-4 (a). Also while polishing, the pad conditioner is activated to keep the pad from being clogged with slurry particles. The pad conditioner is a spinning abrasive disc and the pad conditioning arm moves back and forth on the pad in the direction indicated in Figure 5-4 (a).

Thus, the main variables involved in CMP polishing are the speed of table rotation (table

speed [rpm]), the speed of wafer carrier rotation (quill speed [rpm]), and the force of the polish arm on the wafer (down force [psi]). In addition, jets of air from the wafer carrier push the wafer down. This is called back pressure (back pressure [psi]). It can be used to change the bow of the wafer when polishing in order to change global polish uniformity. For 4-inch wafers, back pressure has a small effect since the wafer is rather stiff. For 6- and especially for 8-inch and greater, back pressure is an important parameter [48].

Many experiments were done on unpatterned wafers with 1 μm thick thermal oxide to determine optimum polishing conditions. It was found that for greatest uniformity on this CMP system, the quill speed should be offset faster than the table speed. Increased down force increases the polish rate, although uniformity suffers with this system when the down force is over 5 psi. Some nonzero back pressure must be applied to keep the wafer from slipping off the wafer carrier. For patterned features things are a bit more complicated. The patterns themselves can interact with the polishing system and have a great effect on polishing results. This is discussed in more detail below.

5.2.2 CMP Mask Design

Global and local CMP polish uniformity is highly dependent on the feature size and distribution, as defined by the patterning mask [49]. This is due to the use of a soft pad which bends around the features when a down force is applied. A harder more abrasive pad can be used to remove features, but this results in poor global uniformity [49]. As shown in Figure 5-5 (a), the spacing between features is important because the soft pad bends into grooves between features and begins to polish into these areas while planarizing the features. And, of the features to be planarized, the smaller features will be removed faster than the larger features since there is more pressure per area on a small feature than a large feature. This results in a global nonuniformity where no matter how much is polished, the areas with larger and/or denser features will remain higher than the areas with smaller and/or sparser features [49]. The amount of nonuniformity can

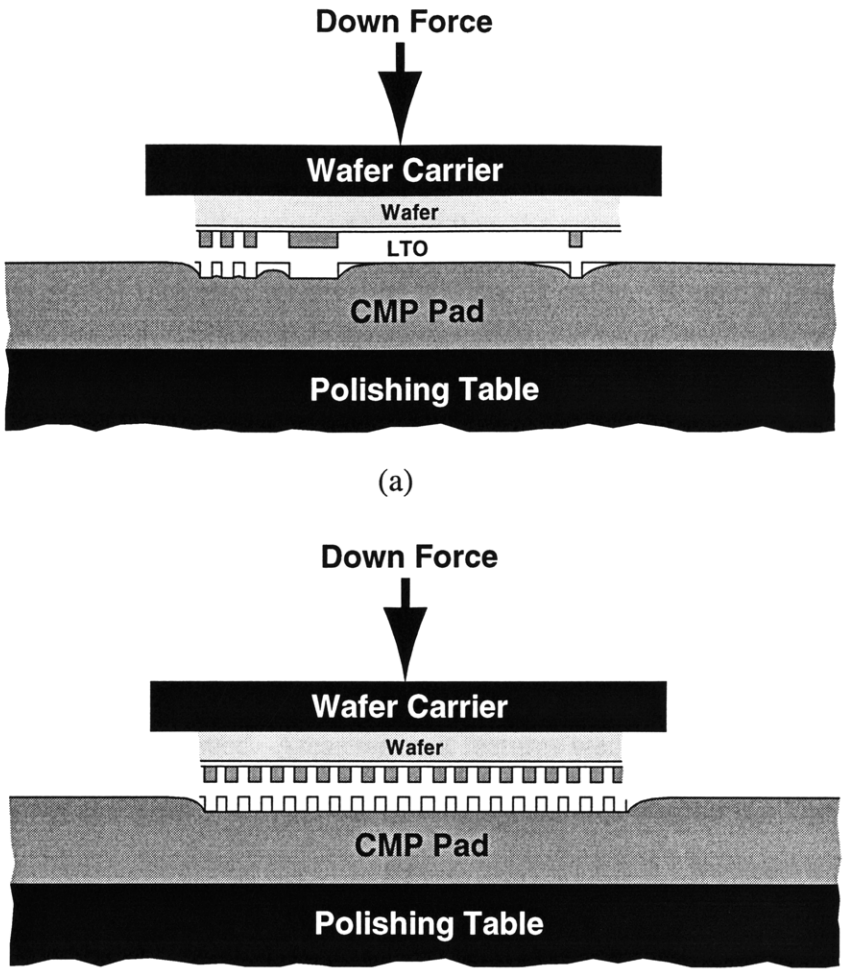


Figure 5-5. Schematic demonstrating CMP pattern dependency when using a "soft" pad. (a) When the features are nonuniformly sized and spaced, down force forces the pad into grooves which. Also, the down force per unit area on the larger features is less than for smaller features resulting in removal of smaller features faster than larger features. (b) These effects are reduced with uniformly spaced and sized patterns.

be decreased with more polishing, but it cannot be removed altogether [49].

Thus, the layout of the features is very important. Better uniformity can be obtained with denser and more uniformly spaced features, as shown schematically in Figure 5-5 (b). Pattern-dependent CMP was a problem in initial bonding experiments with the maskset designed for the double-gate NMOS process. This mask, as shown in Figure 5-6 (a), had dense regions with large features and also sparse regions with small features. The nonuniformity of this mask was

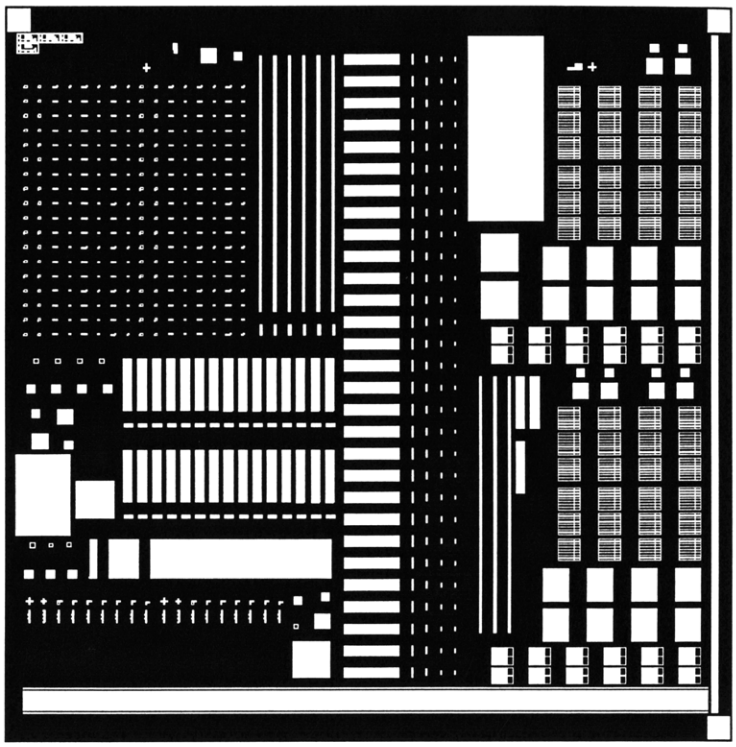
characterized as shown in Figure 5-6 (b)-(e), which show measurements of 12 points in a die, averaged from the center 4 die of the wafer. Nonuniformity is calculated relative to the point in each die with the least amount of oxide remaining after polishing. Areas with the large dense features did not polish as much and remained higher than the areas with the small sparse features, resulting in a 60 nm global step after removal of 800 nm of oxide. Bonding of wafers with this mask was either unsuccessful or showed regions where the transfered film did not adhere after etchback, consistent with the pattern dependence. In comparison, another maskset with more uniformly sized and spaced features was characterized, as shown in Figure 5-7. This mask shows much improved uniformity. Polishing, bonding, and etchback of wafers patterned with this mask was successful.

Thus, the mask in Figure 5-6 (a) was redesigned. The large features which were not necessary for the NMOS process were removed. Areas without features were replaced with dummy lines at 50 % density as shown in Figure 5-8. This use of dummy features is common in current industry practice [46].

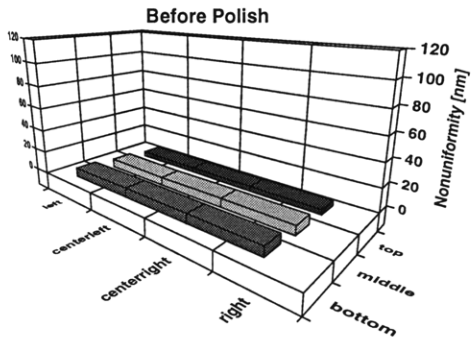
Lastly, due to the "flip", the mask for the flip-bond-transfer process must be the mirror image of the desired image. In order to align to this image after flip-bond-transfer, the flats of the bonded wafers must be well aligned during bonding, and the original image must be well centered and have no rotational skew.

5.2.3 CMP Process

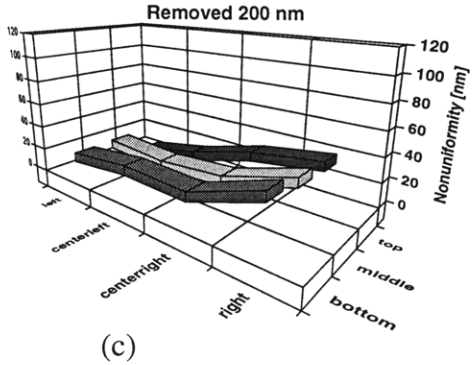
The CMP process was optimized to planarize features resulting form 200 nm thick bottom-gates, as shown in Figure 5-3. The amount of LTO deposited depends on how much needs to be removed in order to planarize the surface. This was dependent on the recipe used, as shown in Figure 5-9, which shows the average step height at 12 points within a die versus how much oxide is removed from areas without features (e.g. scribe lines between dice). Higher speed lower pressure polish is better for uniformity since it minimizes the amount of pad which is bent into grooves



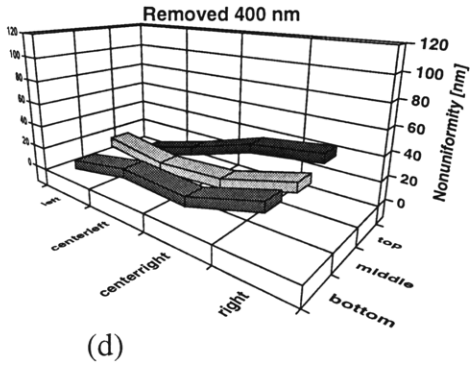
(a)



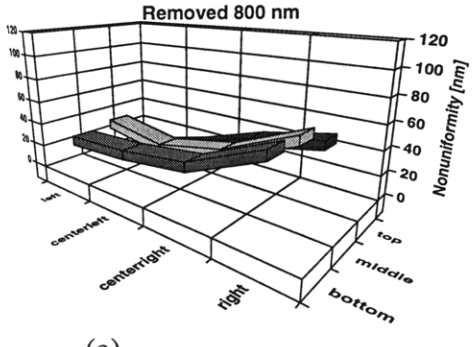
(b)



(c)

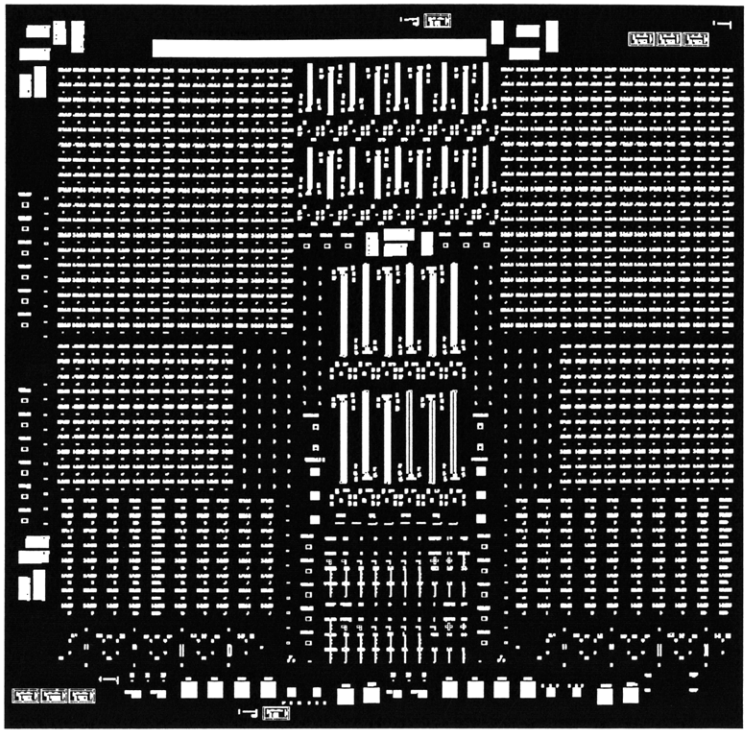


(d)

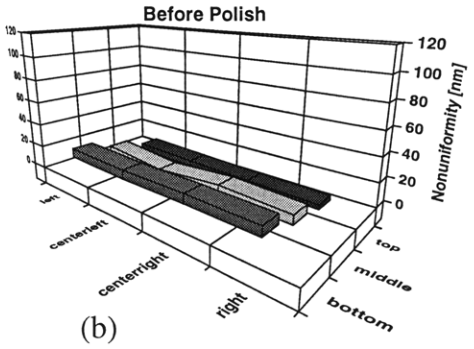


(e)

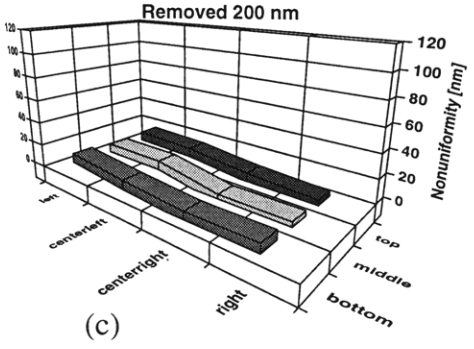
Figure 5-6. Pattern dependency of original bottom-gate mask used for double-gate process. (a) The mask has a great range of feature sizes and spacings. The large features are mainly to the right. (b)-(e) show nonumiformity versus position for 12 points within a die. Nonuniformity is referenced to the point with the thinnest remaining LTO after LTO polish. As LTO is removed, the pattern dependence becomes apparent.



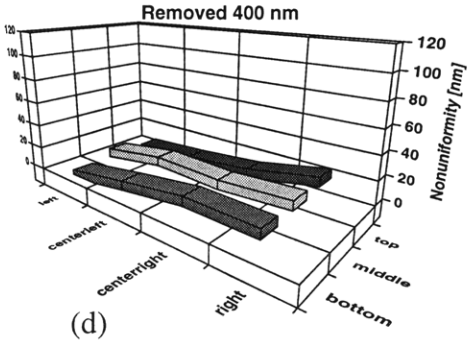
(a)



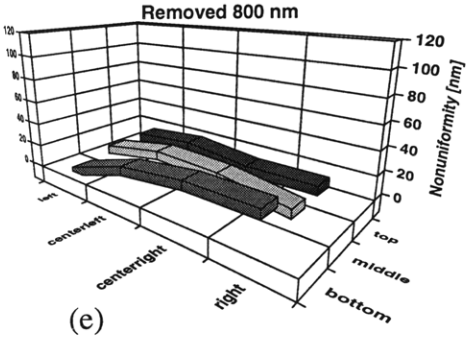
(b)



(c)



(d)



(e)

Figure 5-7. Comparison to results of figure 5-6. Shown is the pattern dependency of a mask from another process. (a) This mask has a much smaller range of feature sizes and more uniform spacing. (b)-(e) show nonumiformity versus position for 12 points within a die. Nonuniformity is referenced to the point with the thinnest remaining LTO after LTO polish. As LTO is removed, the pattern dependence is much less than the results in figure 5-6.

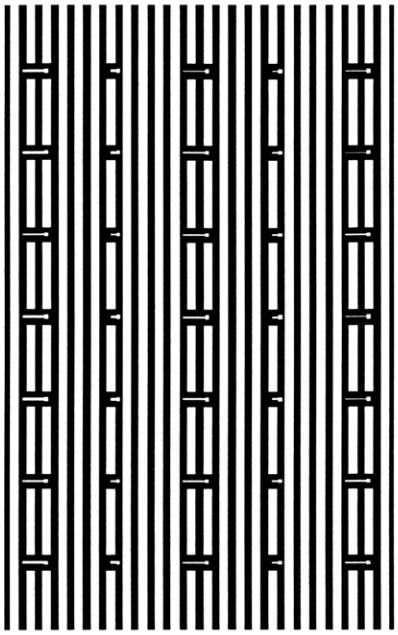


Figure 5-8. A section from final bottom-gate mask optimized for CMP showing the addition of dummy CMP lines in between small features. Large features which are not necessary for the NMOS process were removed and replace with dummy lines.

between features [49]. However, higher speed lower pressure polishing planarizes less than lower speed higher pressure recipes, as shown in Figure 5-9. Thus, for a higher speed lower pressure polish, more LTO must be deposited because more must be polished off to obtain the same planarization as a lower speed higher pressure polish.

It was determined that since the maskset has been optimized for CMP, depositing and polishing less LTO was preferable. The lower speed higher pressure recipe was chosen, and requires removal of about 500 nm of oxide for areas to be planarized, although an additional 400 nm should be removed to ensure local planarization, as shown in Figure 5-10. Thus, for planarization of 200 nm thick bottom-gates, 1.5 μm LTO was deposited and 900 nm to 1 μm was polished away.

After polishing, it is important to remove slurry particles which adhere to the oxide surface during polish. Otherwise, these particles will interfere with wafer bonding. Slurry particles are removed by polishing the wafers without slurry, with only DI water flowing onto the pad. This "water polish" is done in a separate step following the main polish, after the pad has been flushed with DI water for at least 15 minutes to ensure all the slurry particles have been removed. In the

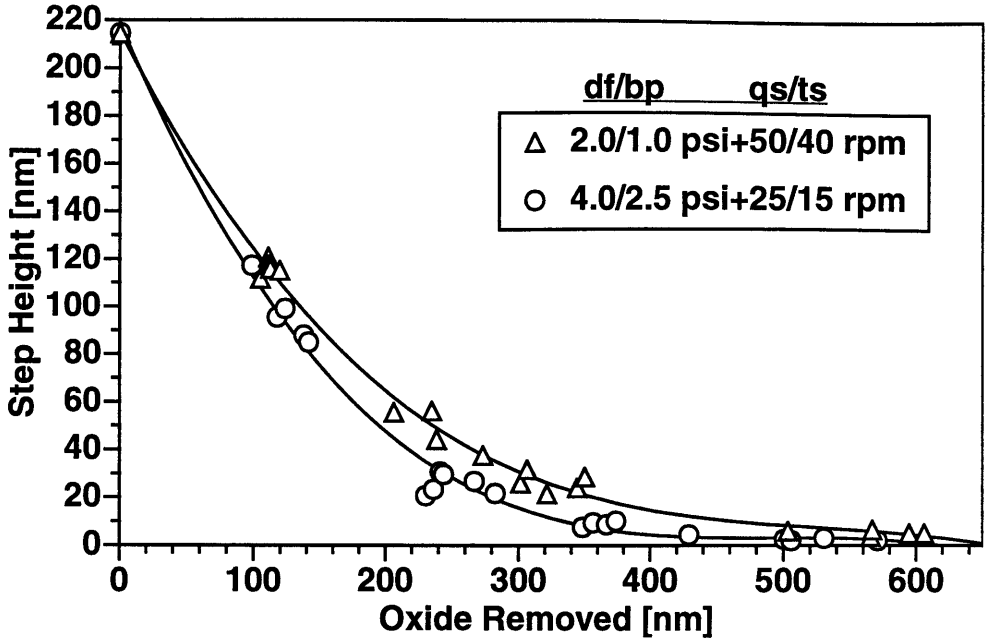


Figure 5-9. Measured CMP planarization characteristics. Shown is the LTO step height versus the amount of oxide removed, which is measured in regions without a step. Two recipes are compared with the down force/back pressure and quill speed/table speed indicated in the legend. In this CMP system, higher down force and lower speeds are better for planarization.

meantime, it is important to keep wafers wet in a tub of DI water because the slurry particles cannot be removed by the water polish if the slurry dries on the wafer (if the slurry does dry on the wafer, more main polish can be done to loosen these particles). This water polish can be optimized to further smooth the surface. As shown in Figure 5-11, 2 psi down force for 40-60 seconds results in a smoother surface.

The CMP recipe is listed in Table 5-1. The polish time is variable. To remove 900 nm of LTO, a time of 390 seconds is typical. Note that this recipe is optimized for planarizing densified LTO. Undensified LTO polishes at a faster rate and planarizes with different characteristics with this recipe. The LTO used in the process was densified at 1000 °C for 10 minutes. Other temperatures and times can be used for densification, so long as the same level of densification is achieved. The method used to measure whether temperature and time combinations lead to densification is measurement of wet etch rate in BOE. The LTO etch rate is approximately 400 nm/min without densification, decreases with increasing densification temperature and time to a minimum of

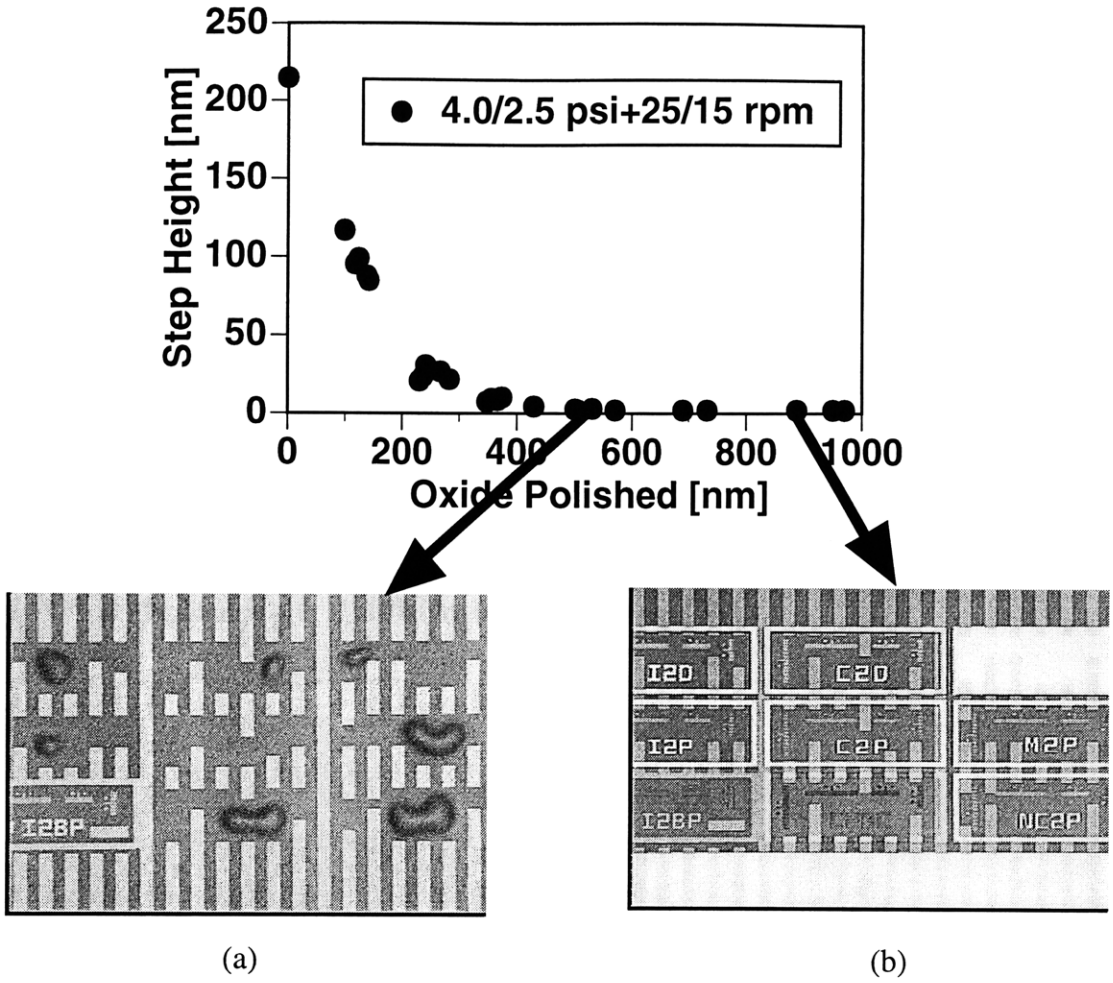


Figure 5-10. Schematic showing the need to overpolish. (a) Shows a wafer polished to where the step height is unmeasureable. However, small voids form in regions where dummy CMP lines were discontinuous. (b) This sample was over-polished and shows no voids.

approximately 160 nm/min.

After the main polish and water polish, the wafers are cleaned in an Evergreen wafer leaning system. This machine rinses the wafers with ammonia and DI water. After this, a piranah clean is performed prior to reintroduction into the cleanroom. Prior to wafer bonding, the wafers are run through a full RCA clean.

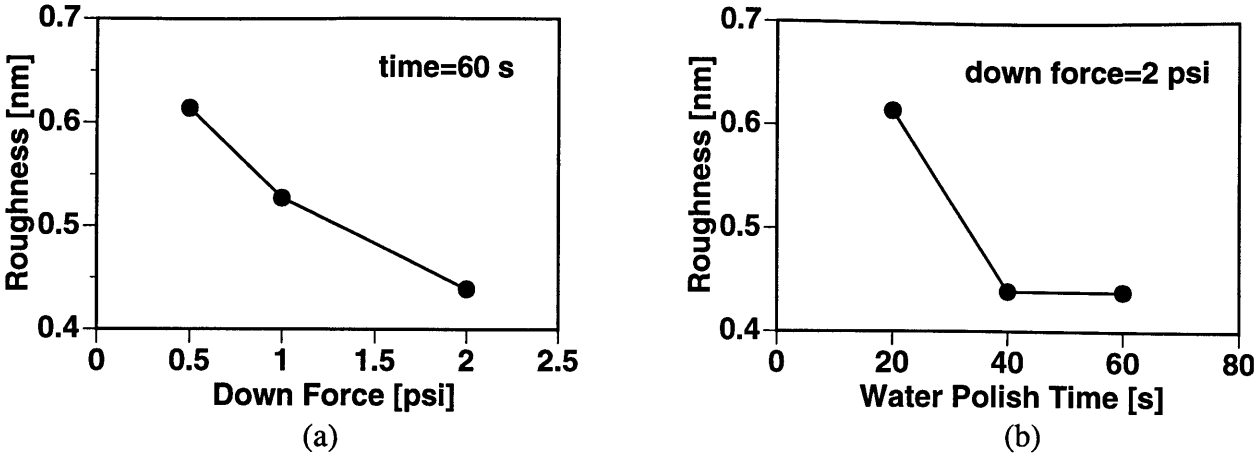


Figure 5-11. AFM measurement of LTO surface roughness after water polish versus (a) down force, and (b) time. These measurements were performed on 1.5 μm LTO which had been densified at 1000 °C for 10 minutes and CMP-ed to remove 500 nm LTO.

	Main Polish	Water Polish
Time [s]	variable	60
Down Force [psi]	4.0	2.0
Back Pressure [psi]	2.5	0
Quill Speed [rpm]	25	60
Table Speed [rpm]	15	60
Slurry [ml/min]	150	0

Table 5-1. CMP recipe used for planarization. Note that the pad conditioner is used in-situ during the main polish. The pad must be flushed out with DI water to remove slurry from the pad. Typically, the main polish is performed on a set of wafers which are kept in a DI water bucket after they are polished. The pad is flushed with running DI water for at least 15 minutes, then each wafer is water polished.

5.2.4 Wafer Bonding

Bonding phenomena between two flat smooth surfaces have been known for a long time. The first widespread use with silicon wafers was for fabrication of SOI wafers with the bond-etchback BESOI technology as discussed in chapter 1. Since then, it has been used extensively for fabrication of micro-electro-mechanical devices and systems (MEMS) [50].

Bonding of two cleaned, smooth, flat, bare, prime-grade silicon wafers can occur via two mechanisms. The first results in an oxide-oxide bond, and the second results in a silicon-silicon bond. An oxide-oxide bond occurs because a native oxide exists on bare silicon wafers or is formed during cleaning. The cleaning process saturates the surface with several monolayers of hydroxyl groups (OH^-), resulting in a hydrophilic surface. When two wafers are brought together, these hydroxyl groups are attracted to each other by van der Waals forces and hence form a weak bond at room temperature. At elevated temperatures, covalent bonds are formed between the two wafers by the reaction $\equiv Si\text{-}OH^- + OH^-\text{-}Si\equiv \rightarrow \equiv Si\text{-}O\text{-}Si\equiv + H_2O$, where "-" denotes a covalent bond and "≡" denotes three satisfied covalent bonds of the silicon surface atoms [47].

A silicon-silicon bond is formed if the wafers are stripped of oxide by HF treatment after the clean step. This leaves the surface silicon atoms terminated with hydrogen. However, exposure to air allows some hydroxyl groups and hydrocarbons to be absorbed. It has been proposed that these are responsible for the initial van der Waals bonding at room temperature between the two otherwise hydrophobic wafers at room temperature [47]. However, the bond has been measured to be at least 2-3 times weaker than a room temperature bond between hydrophilic wafers [47]. At elevated temperatures, direct Si-Si bonding occurs and elevates the bond strength of the bonded pair to the silicon fracture strength. A plot of surface energy versus anneal temperature is reproduced from [47] and shown in Figure 5-12. Silicon-silicon bonding produces a much stronger bond at lower temperatures than does oxide-oxide bonding.

Regardless of the bonding mechanism, the actual quality of the wafer-wafer bond depends greatly on surface smoothness, surface treatment, bonding procedure, anneal temperature, and anneal

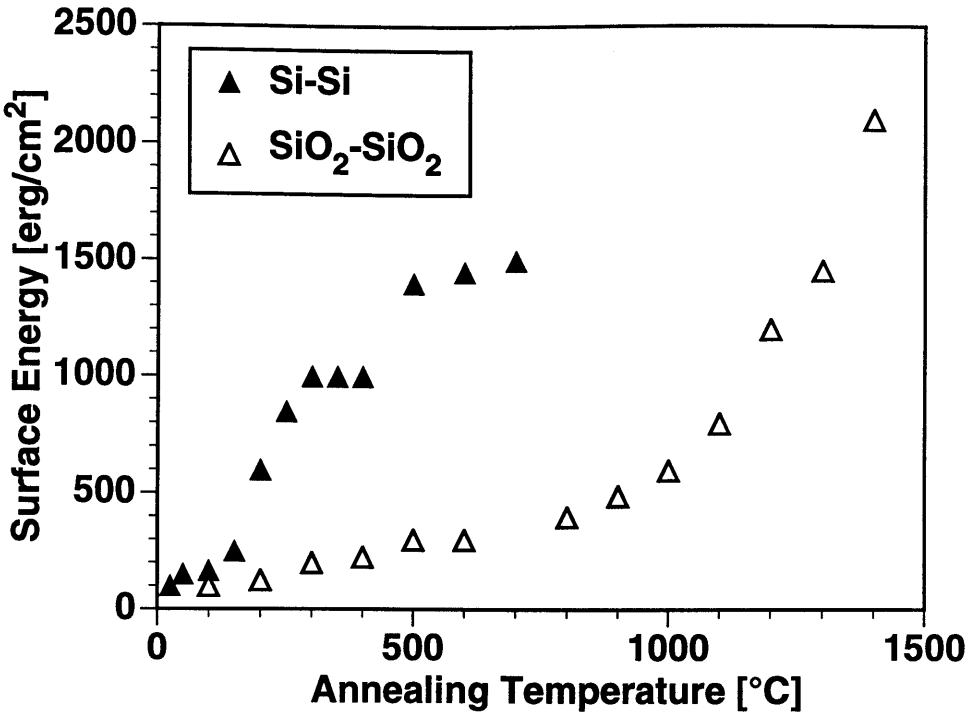


Figure 5-12. Bond strength versus anneal temperature for hydrophilic SiO_2-SiO_2 bonding, and for hydrophobic Si-Si bonding, reproduced from [47].

time. Recall that the bond of interest is between polished densified LTO and thermal oxide. The surface roughness of the LTO through the CMP process, and of the thermal oxide is listed in Table 5-2. The wafers are bonded with the jig shown in Figure 5-13. This bonding jig consists of 3/4 inch thick teflon board with two flat alignment bars at the bottom (one fixed, one removable). Three 2x1/4 inch by 7 mil thick teflon shims are used to keep the handle wafer and SOI wafer apart before bonding. A teflon alignment guide, shown near the top of 5-13 (a) is used to push the handle wafer and SOI wafer against the alignment bars at the bottom of the board for flat-flat alignment. A teflon roller, shown in Figure 5-13 (b), is used to make the initial contact of the wafers in the center, and to roll out any voids due to trapped air.

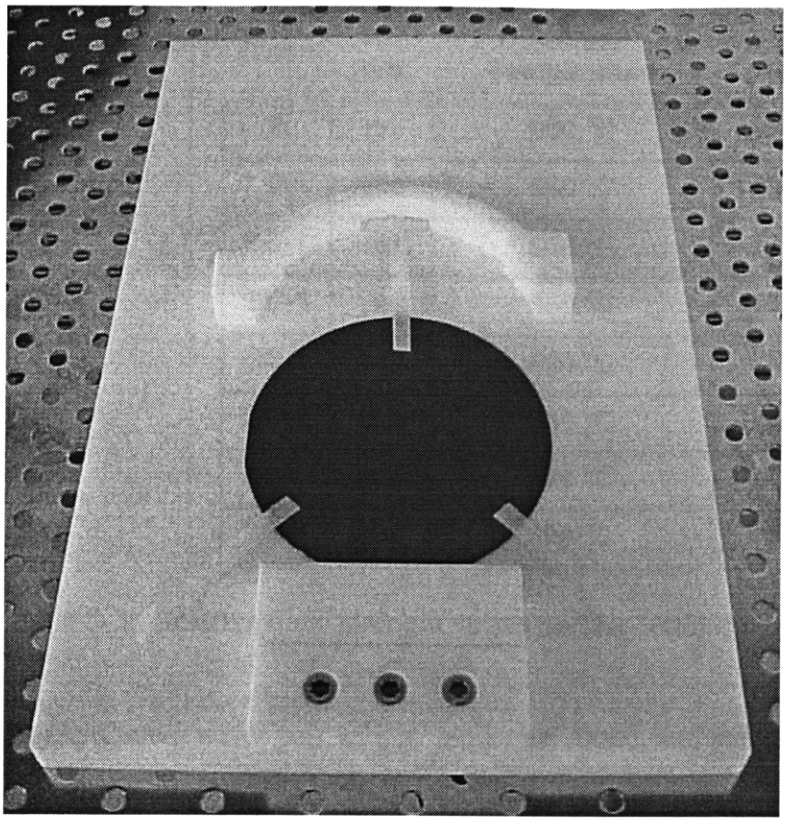
The optimum cleaning procedure for this process is as follows: 1) 10 minute SC-1 clean, 2) full dump rinse, leave the DI water running, 3) 20 second 50:1 HF dip, 4) 13 second rinse in the running DI water, and 5) finally a 45 second spin dry with the spin dryer set at 3000 rpm (note, this does not mean 45 seconds at 3000 rpm, just a total of 45 seconds -- it is important not to overdry

	Roughness [nm]
Bare Si (p-prime)	0.217
Thermal Oxide (100 nm)	0.253
LTO (120 nm)	1.094
LTO (1.5 μm)	5.202
LTO after main polish†	0.531
LTO after water polish	0.439

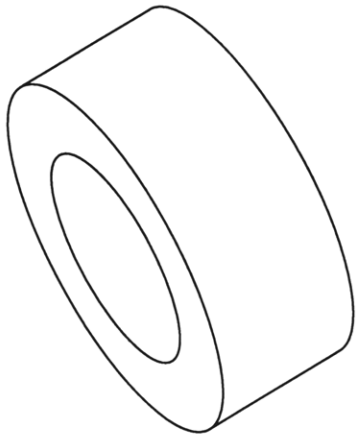
Table 5-2. AFM measurement of surface roughness. A bare p-prime wafer is used as reference. †LTO after main polish surface roughness does not include imbedded slurry particles, which were measured to be approximately 200 to 700 nm in size.

wafers -- running a full 5 minute spin dry did not produce good bonding results). Typically, two wafer pairs are cleaned and bonded at a time. The handle wafer is placed face up on the bonding jig with the flat on the alignment bar at the bottom of the board. The three shims are put in place as shown in Figure 5-13 (a). The SOI wafer is placed face down the three shims. The top alignment guide is pushed down on the two wafers so that both are flush to the bottom alignment bar. At this point, the roller is pushed on the middle of the wafers and the shims are removed as quickly as possible. Keeping the roller depressed, the alignment guide and bar are removed and manual pressure in a slight rocking motion is applied on the roller to make sure there is sufficient contact between the two wafers. The bonded wafer pair is then inspected with an IR inspection system for voids [50]. The wafer pairs typically are void-free, but in the event of a void due to trapped air (irregularly shaped voids), the roller can be used to press it out. Voids due to particles (voids shaped like concentric circles) cannot be rolled out. These wafer pairs must be split apart with the edge of a teflon tweezer while immersed in water, then recleaned before rebonding.

The anneal temperature, anneal time, corresponding surface clean and HF dip, and relative bond strength are listed for several processes in Table 5-3. This is a tabulation of results for a series of bonding experiments undertaken with both thermal-thermal and thermal-polished-LTO wafer pairs to 1) find the optimum bonding procedure and 2) to see how low in temperature wafers can be



(a)



(b)

Figure 5-13. (a) Wafer bonding jig shown with a handle wafer, 3 shims, and flat alignment guides. The SOI wafer is placed face down onto the shims. The alignment guides align the flat. The roller shown in (b) is used to make contact at the center of the wafer, at which time the 3 shims are removed.

well bonded. Note that the polished-LTO wafers have dummy features under the LTO consisting of patterned 200 nm thick polysilicon on 500 nm thick thermal oxide. Each tabulated process is a result of at least three separate bonding experiments, each with at least two pairs of each type of bond. The thermal-thermal and thermal-polished-LTO bonds did not differ much, although the thermal-polished-LTO tended to chip a little more when a "strong" bond was present.

Bond strength was measured by inserting a razor blade between the two wafers of the bonded pair [47]. This is repeated around the entire perimeter of the bonded pair. The razor blade cannot be inserted into a "strong" bond without immediately chipping the wafers. Chips which fall off are sometimes discolored, indicating a transfer of an oxide layer from one wafer to another. When these bonded pairs are shattered, they break as if they were one wafer. The razor blade can be

Clean	HF dip	Anneal Temperature	Anneal Time	Bond Strength
P	no	950 °C	40 min	weak
P	yes	950 °C	40 min	moderate
SC-1	no	950 °C	40 min	moderate
SC-1	yes	950 °C	40 min	strong
SC-1	yes	800 °C	1 hr	moderate
SC-1	yes	800 °C	2 hr	strong
SC-1	yes	600 °C	1 hr	weak
SC-1	yes	600 °C	4 hr	strong

Table 5-3. Relative bond strength versus surface treatment and thermal anneal in N_2 ambient.

inserted into a wafer pair with a "moderate" bond, and pieces of one wafer can be broken off with twists of the razor blade. However, pieces toward the center stay adhered. The pieces which do break off are not discolored. Finally, the razor blade can be inserted into a wafer pair with a "weak" bond, and when worked around the perimeter, the razor blade can be used to completely separate the two wafers intact.

From Table 5-3, note that the SC-1 clean is better than piranah clean. The SC-1 clean removes more particulates than the piranah clean because it etches the surface of the wafer during the clean, particularly at high temperatures near or at 80 °C, whereas the piranah clean does not [51]. The HF dip also improves the oxide-oxide bond strength. This result is previously unreported and was learned by personal correspondence with staff at Lincoln Laboratories [52]. Dilute solutions of HF are known to further remove any surface particulates [51]., and this may explain the difference in oxide-oxide bond strength with and without HF. However, it is suspected that a different surface chemistry may occur because this process can be pushed down to 600 °C with results similar to a silicon-silicon bond. However, without quantitate bond strength measurements and rigorous surface analysis, it is not possible to know the exact nature of this observed improvement in bond strength.

In any case, reliable polished-LTO to thermal oxide bonds were achieved at anneal

temperatures down to 600 °C. This pushes the minimum process temperature down to 650 °C (recall that LTO should be densified at least 50 °C above the bonding temperature and that all subsequent processing be 50 °C less than the densification temperature). This is important if the flip-bond-transfer process is to be used for back-end applications, as will be described in chapter 6.

To test whether these bonds can produce the desired film transfer, the test wafers with polished-LTO on patterned dummy poly gates were subjected to chemical etchback with 10% KOH solution at 90 °C using the 500 nm oxide as an etchstop [53]. "Strong" and "moderate" bonds survived the substrate removal process, but "weak" bonds do not - the transfered LTO layer does not adhere to the handle wafer after etchback.

5.2.5 Substrate Removal

The substrate removal process with KOH is not CMOS compatible due to ionic contamination. A more compatible etchback chemistry is etching with in tetramethyl ammonium hydroxide, TMAH, 25 % weight in water, at 80 °C. TMAH is CMOS compatible. In dilute form it is used as a photoresist developer. At 80 °C, it etches silicon at a rate of 23 μm/hour with excellent selectivity to thermal oxide 9,200:1, deposited oxide 2950:1, and silicon nitride 33,000:1 [54]. Because the etch rate is rather slow, mechanical grinding is used to remove most of the substrate before TMAH etch. With 4-inch wafers which are 525 to 550 μm thick, 450 μm are removed by grinding. The nonuniformity of the mechanical grinding can be tolerated because the TMAH stops well on the buried oxide of the SOI wafer.

Recall that the handle wafer has 500 nm thermal oxide. This should protect the handle wafer from TMAH during etchback. Figure 5-14 shows a successfully transfered SOI layer on the handle wafer. However, the edges were attacked by the TMAH. This is perhaps due to the poor quality thermal oxide which grows on the edge of wafers. To fix this problem, the etchback process was modified as shown in Figure 5-15. Immediately after bond anneal, the wafer pairs are coated with 120 nm LPCVD nitride. This is done twice (for a total of 240 nm), where on the second run

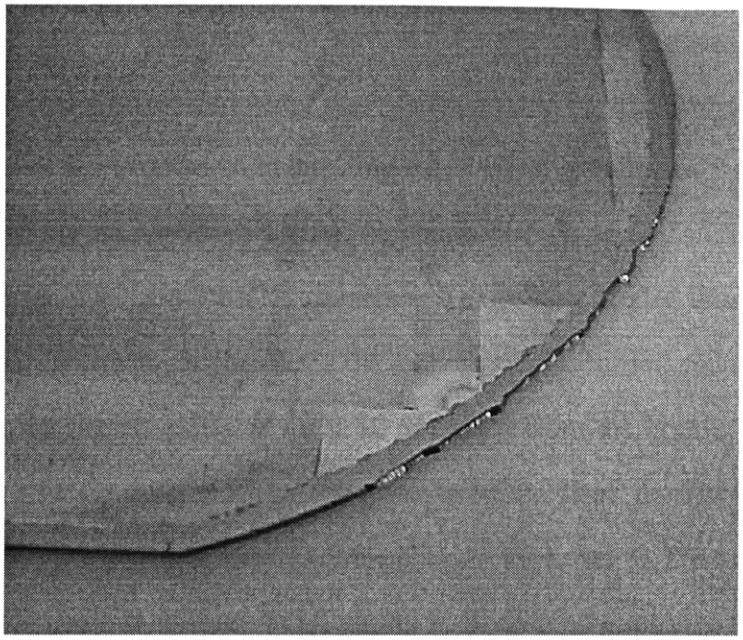


Figure 5-14. Photograph of a successfully delaminated SOI layer. However, the edge of the a wafer was attacked by TMAH etch. The handle wafer thermal oxide at the edge is not a sufficient mask against TMAH.

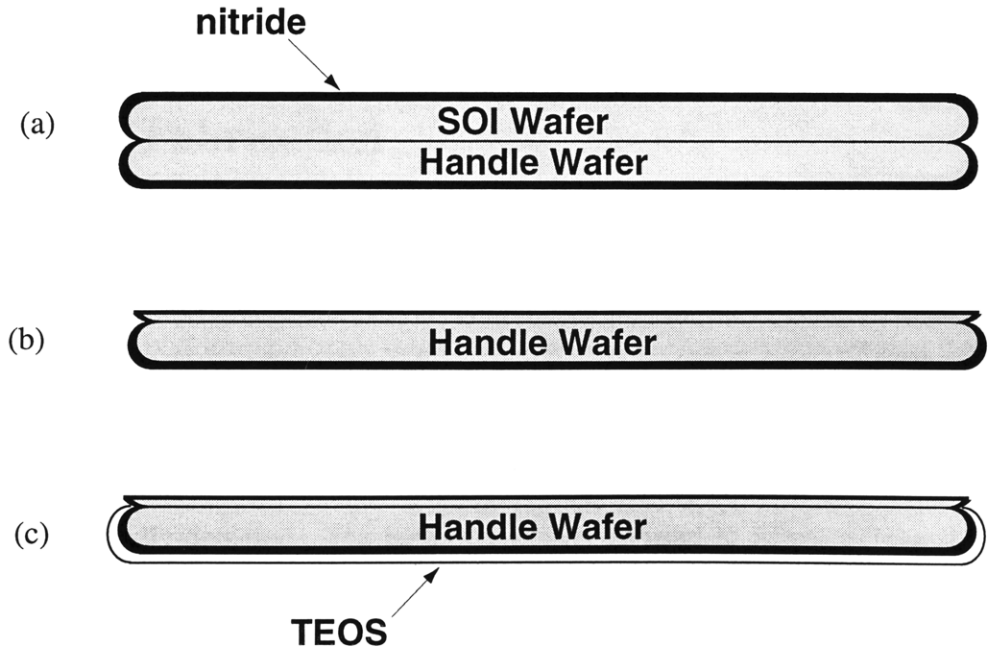


Figure 5-15. Method to protect the edge of the wafer from TMAH. (a) After bond anneal, the wafer pair is put in LPCVD nitride. (b) The SOI wafer substrate is ground back. (c) An additional layer of PECVD TEOS is deposited just in case the nitride is damaged during grindback.

the wafers are rotated 180 ° to make sure that parts of the wafer sitting in the rails also receive a coat of LPCVD nitride. The nitride-sealed wafer pair is then ground back. To insure that the edges and back of the wafer have not been scratched during mechanical grinding, a 5 μm layer of PECVD TEOS is deposited on the back prior to TMAH etchback.

If the SOI wafers used are SIMOX, there is one extra step needed during TMAH etchback. SIMOX wafers have silicon inclusions near the bottom of the buried oxide layer [55]. These islands of silicon in the buried oxide, at most 50 nm away from the bottom of the buried oxide. These may interfere with removal of the buried oxide in BOE. Thus, the TMAH etch process is as follows: 1) 2 second BOE dip to remove native oxide, 2) DI rinse, 3) TMAH etch at 80 °C until wafer clears and SOI layer is visible, 4) DI rinse, 5) BOE dip to remove 50 nm buried oxide, enough to expose silicon inclusions, 6) DI rinse, 7) 30 second TMAH etch to remove silicon inclusions, and 8) DI rinse. When not using SIMOX wafers, steps 5) through 8) should be omitted. Following TMAH etchback, the wafers are cleaned in piranah before reintroduction into the cleanroom, at which point the remainder of the buried oxide is removed in BOE.

5.3 Device Fabrication

For demonstration, double-gate NMOSFET's were fabricated on successfully transfered SOI layers. The silicon film thickness for these NMOSFET's was 45 nm, the top-gate oxide thickness was 6 nm, and the bottom-gate oxide thickness was 20 nm. Gate lengths ranging from 5 μm down to an L_{eff} of 0.24 μm were fabricated. The gate transfer *I-V* characteristics of the top- and bottom-gate are shown in Figure 5-16 (b). The kink effect present in the top-gate *I-V*'s indicate that the devices are partially-depleted. The bottom-gate can be used to adjust the threshold voltage, as shown in Figure 5-16 (c), and can be used to change the device from partially- to fully-depleted. This is shown in Figure 5-16 (d), where application of the bottom-gate voltage removes the kink in output characteristic *I-V* curves.

A split of these devices also received a bottom-gate counterdoping implant through the top-

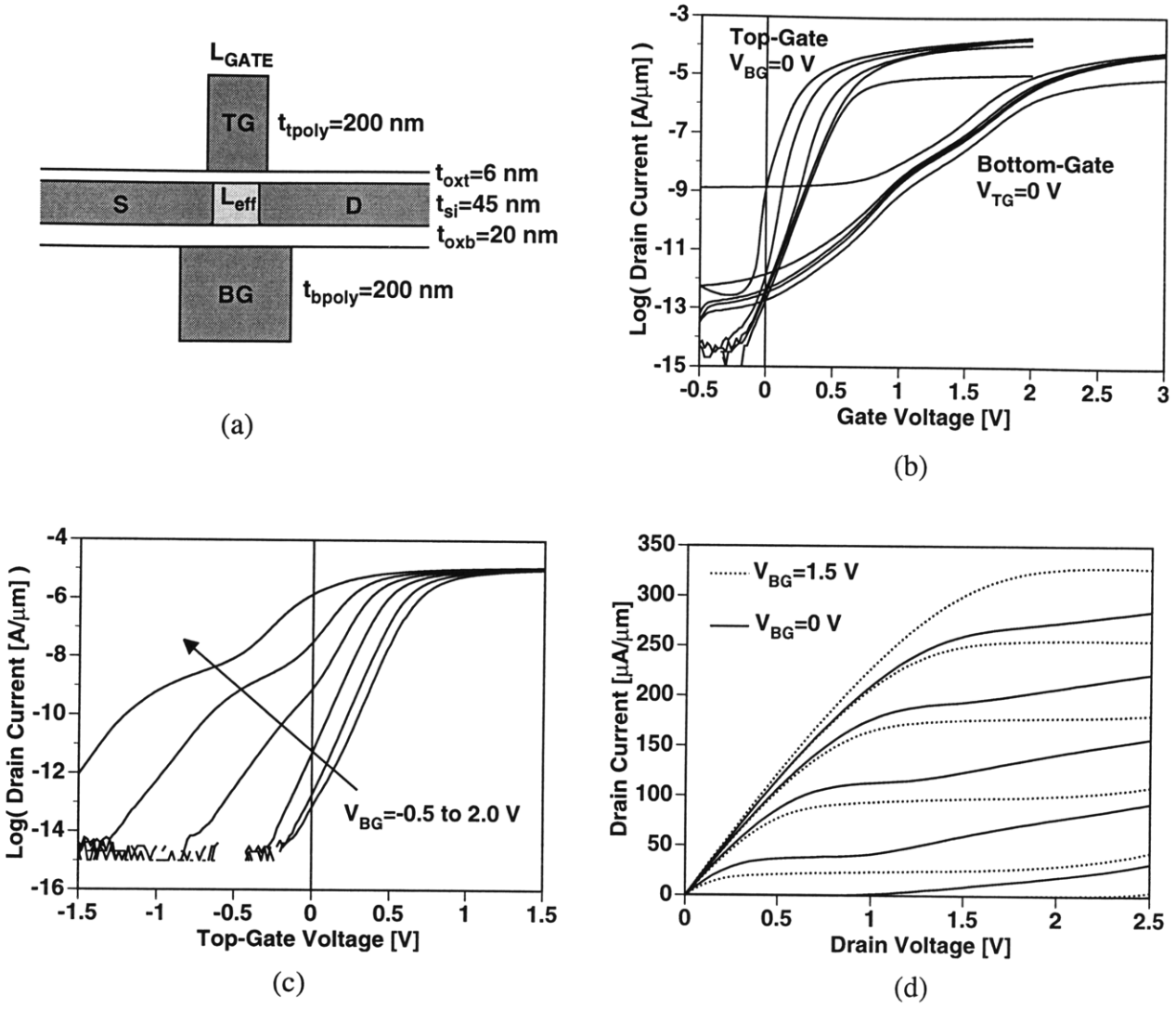


Figure 5-16. (a) Thickness dimensions of fabricated devices. The following plots are for a device with L_{eff}=0.24 μm. (b) Top-gate and bottom-gate *I-V* characteristics. Each is measured with the other gate at 0 V. Each curve represents V_{DS}=0.05, 0.55, 1.05, 1.55, and 2.05 V. (c) Top-Gate *I-V* chacteristic for V_{DS}=0.05 V versus bottom-gate voltage from -0.5, 0, 0.5, 1.0, 1.5, and 2.0 V. (d) Top-gate output characteristics with bottom-gate voltage at 0 and 1.5 V. Each curve represents V_{GS}=0.5, 1.0, 1.5, 2.0, and 2.5 V.

gate. These devices did not work. It appears as if the top-gate poly was not a sufficiently thick masking layer to prevent the tail of the counterdoping implant from doping the channel. The source and drain was shorted out by a n-type channel. This problem can be corrected for NMOSFET's by using insitu-doped polysilicon and etching it with a tall oxide hardmask, which makes the top-gate stack a better mask against the counterdoping implant. In reality, the specifics of this problem are

not important because the thickness of the layers in this experiment are not typical of scaled devices. A study of counterdoping for scaled dimensions is presented below.

5.4 Top-Bottom-Gate Self-Alignment

Counterdoping the bottom-gate reduces the source/drain to bottom-gate parasitic capacitance. This also minimizes any layout-dependent parasitic capacitances which can occur if the top-gate is misaligned to the bottom-gate. In effect, self-alignment is achieved if the counterdoping is effective. Figure 5-17 schematically compares the source/drain to bottom-gate overlap capacitance for various doping configurations of the bottom-gate. In Figure 5-17 (a), the bottom gate has not been counterdoped, thus the capacitance is high, as indicated by the close proximity of the dashed lines, which represent depletion region edges on either side of the capacitance. If the bottom-gate region underneath the source/drain is counterdoped and the doping is lowered by several orders of magnitude, then the capacitance can decrease significantly as shown in Figure 5-17 (b). If the region underneath the source/drain is counterdoped n-type, it is then electrically floating in the p-type bottom-gate. When the source/drain voltage changes, this region is capacitively coupled to the source/drain voltage. Therefore, the resulting source/drain to bottom-gate capacitance is between the n-type counterdoped region and the rest of the p-type bottom-gate, as shown in Figure 5-17 (c). So long as the counterdoped region is not too high in concentration, this should still be effective in reducing capacitance. However, counterdoping and producing a highly doped n-type region results in high capacitance, as shown in Figure 5-17 (d).

Thus, an effective counterdoping implant is not an easy task, mainly because the bottom-gate should be highly doped in order to minimize gate resistance. It can be silicided, but if the bottom-gate doping is too low, the polysilicon will deplete, reducing bottom-gate control. The problem with high doping is that it is difficult to bring the doping level down by orders of magnitude. For example, reducing a doping concentration from $1x10^{19}$ cm^{-3} to $1x10^{16}$ cm^{-3} requires a change of $9.999x10^{18}$ cm^{-3}. Such accuracy is possible, as shown in Figure 5-18, which shows SUPREM4

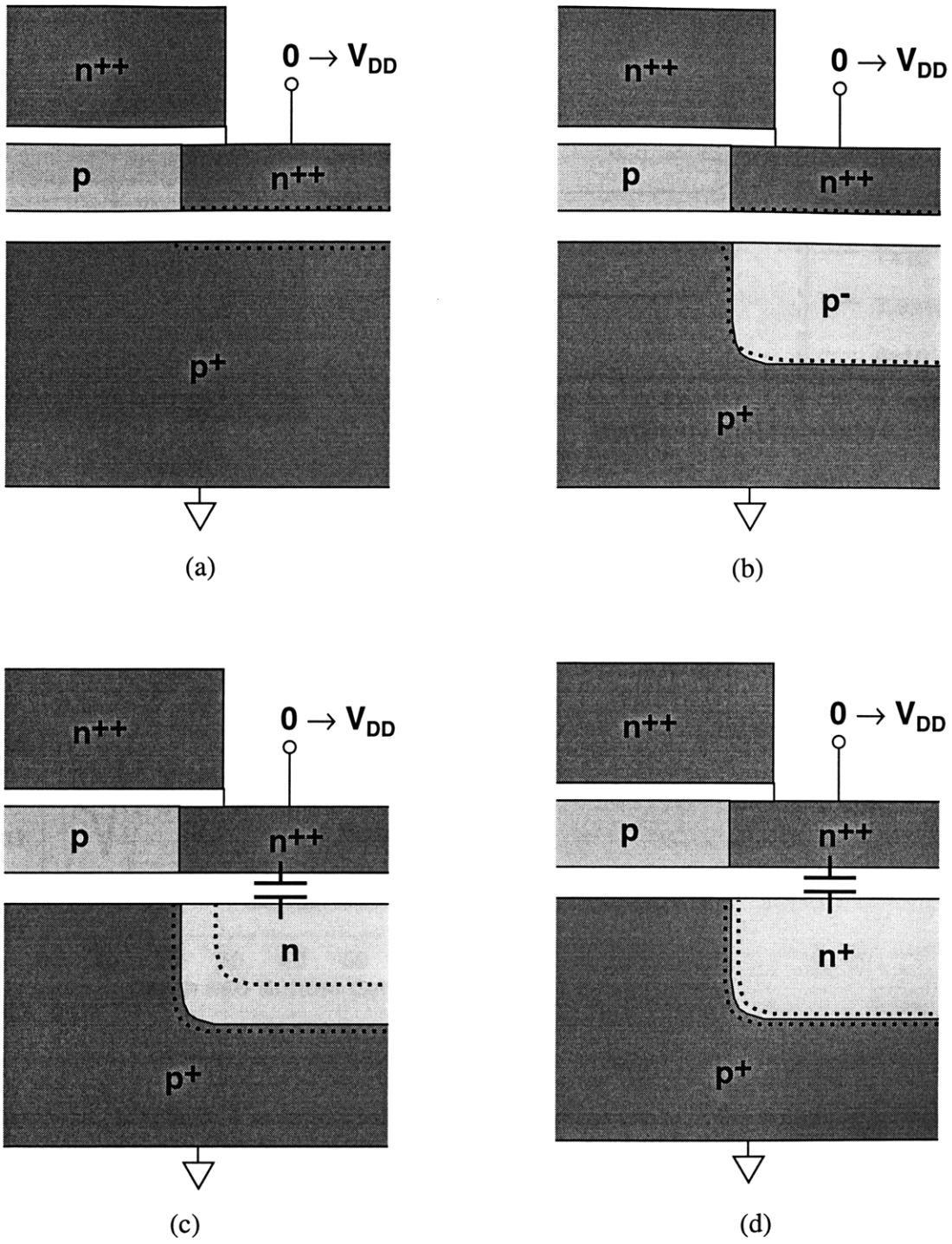


Figure 5-17. Schematics of source/drain to bottom-gate overlap demonstrating the effectiveness of counterdoping. The dashed lines represent depletion region edges. The farther apart the depletion region edges, the smaller the capacitance. (a) Without counterdoping the parasitic source/drain-bottom-gate capacitance is large. (b) Ideal reduction of capacitance occurs when the overlapping region is counterdoped p^-. (c) If the counterdope region becomes n, it is capacitively coupled to the source/drain. Thus, the capacitance is between the p-n region in the bottom-gate. (d) If the counterdope is heavy n^+, the capacitance is as bad or worse than no counterdoping.

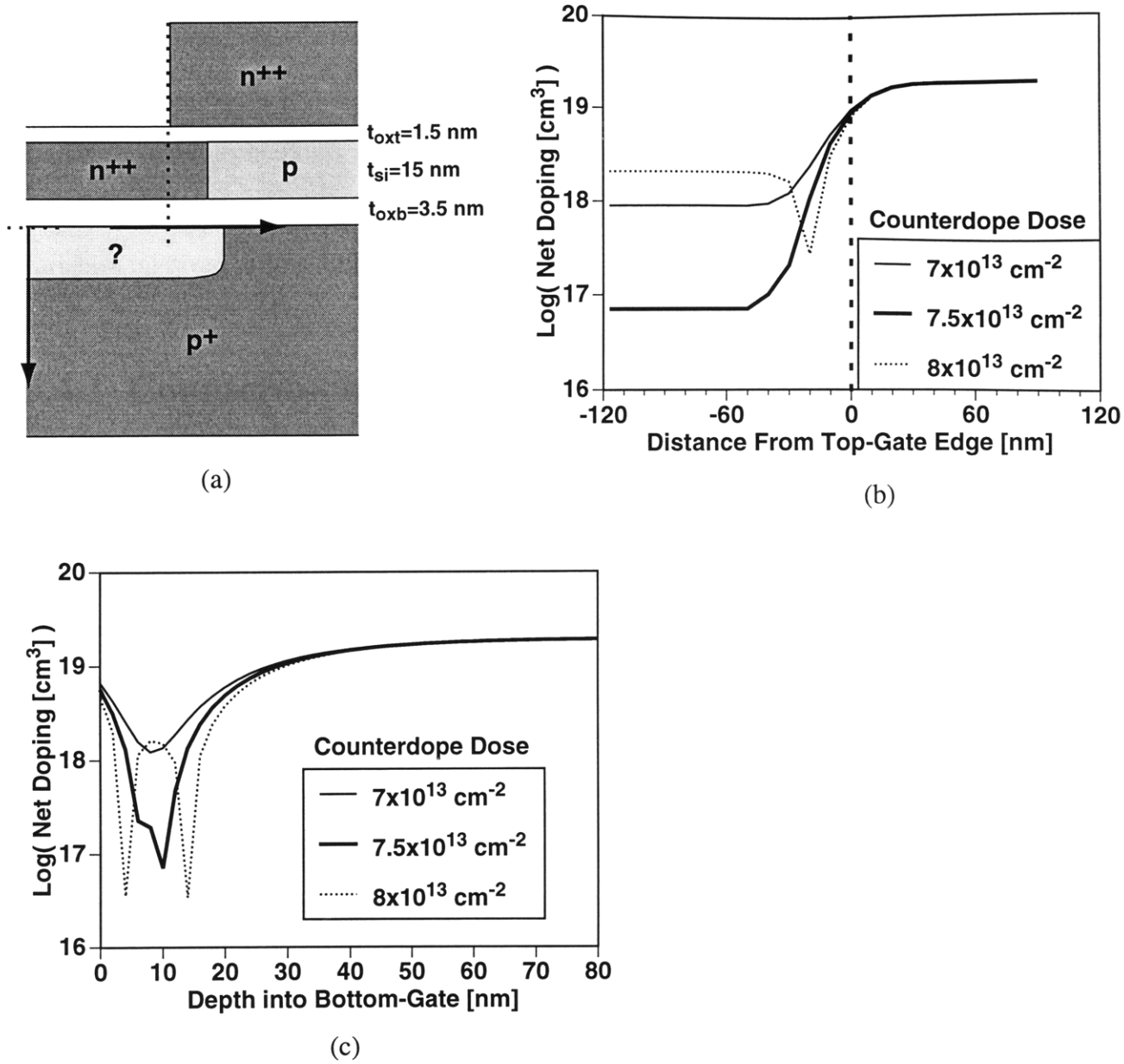


Figure 5-18. SUPREM4 as-implanted profiles in the structure with the bottom-gate doped $2x10^{19}$ cm^{-3}, as shown in (a). (b) shows the counterdoped lateral doping profile referenced to the top-gate edge. Note the lateral straggle of the counterdoping implant. (c) shows the counterdoped profile into the bottom-gate.

simulation of as-implanted doping profiles in a structure with thickness typical of scaled fully-depleted SOI MOSFET's [24].

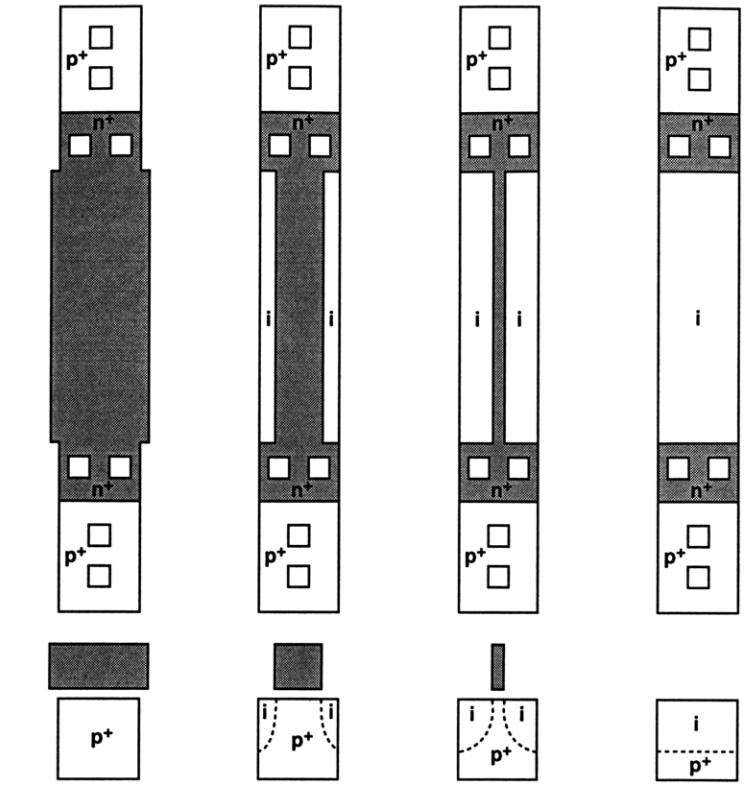
However, counterdoping may be impractical given the optimization required and the possible sensitivities to process variations. Actual doping activation depends highly on implanted species,

implant conditions, process temperatures and times, and the polysilicon material used [45]. Also of concern is limiting the lateral subdiffusion of the counterdoping implant, since this will limit the scalability of this technology. As shown in Figure 5-18 (b), the as-implanted lateral straggle already encroaches the top-gate edge by 30 nm. The following experiment was performed to assess the variables involved in counterdoping the bottom-gate for self-alignment.

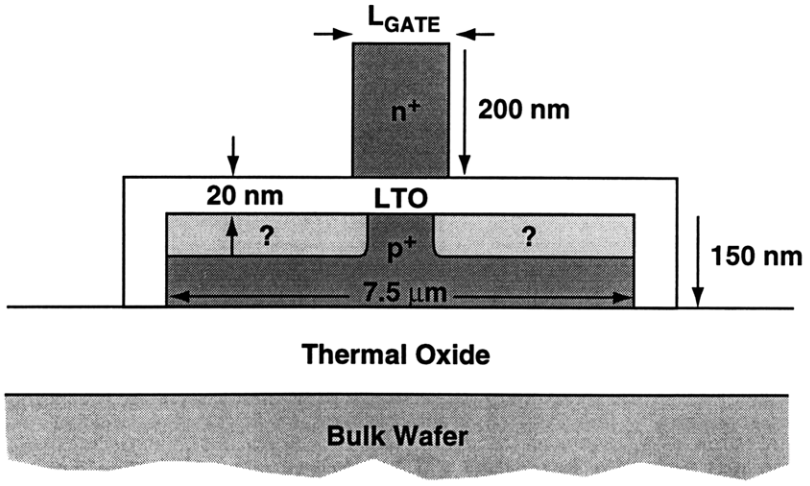
5.4.1 Counterdoping Experiment

Counterdoping the bottom-gate is made difficult by the need reduce the bottom-gate doping concentration with an implanted doping profile through thin layers. The resulting as-implanted profile is sharply peaked, due to the relatively small implant range [45]. Ideally, this should be annealed until the profile is flat, but this results in lateral subdiffusion which reduces bottom-gate control. This experiment clarifies this point by using a low-temperature 600 °C anneal to reveal the approximate as-implanted profile. These profiles are characterized electrically in the resistive test structures shown in Figure 5-19. Figure 5-19 (a) shows that an array of bottom-gates are implanted through various top-gate lengths. This results in a variable resistance which can be used to separate lower-doped counterdoped regions from the highly-doped bottom-gate regions.

A simplified process is employed to fabricate these test structures. A bulk wafer is used. Oxide is grown, and then the bottom-gate material is deposited. A split was done were both amorphous-silicon was deposited at 560 °C and fine-grain polysilicon was deposited at 600 °C. The bottom-gate is implanted with boron, patterned, and etched. A 20 nm LTO layer was deposited to mimic the device layer of scaled fully-depleted SOI MOSFET's. The wafers then receive a 14 hour recrystallization anneal at 600 °C. This transforms the amorphous silicon into large-grained polysilicon, but does not affect the grain size of the fine-grain polysilicon [45]. A top-gate polysilicon is then deposited, patterned, and etched, followed by an arsenic counterdoping implant. The final dimensions of the test structures is shown in Figure 5-19 (b). L_{GATE}'s down to 0.2 μm, as measured in SEM, were fabricated.



(a)



(b)

Figure 5-19. (a) Resistor array used to experimentally characterize counterdoping implants. (b) Dimensions of fabricated resistor arrays. The shortest fabricated L_{GATE} was 0.2 μm.

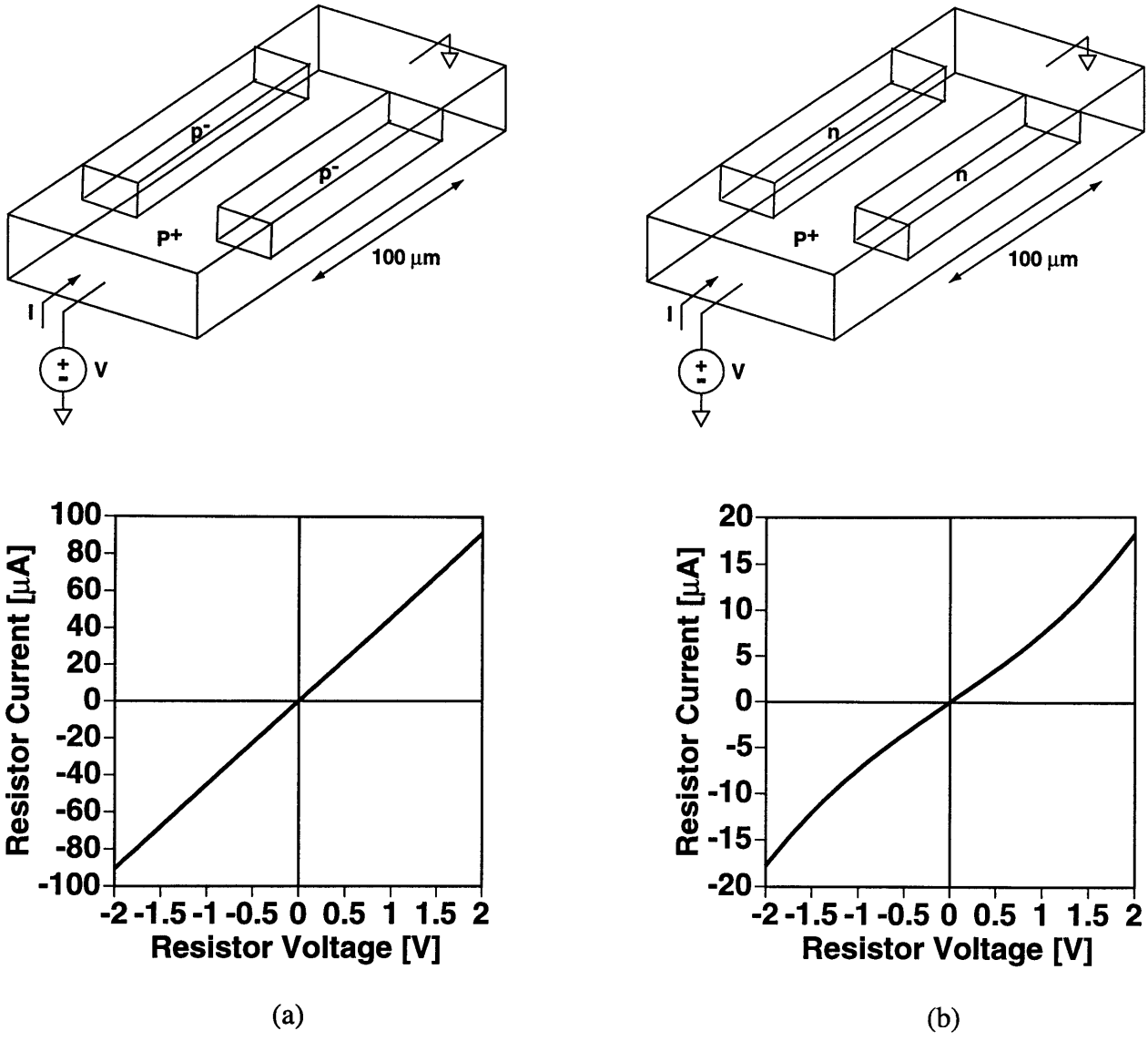
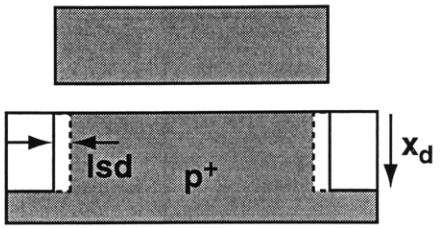
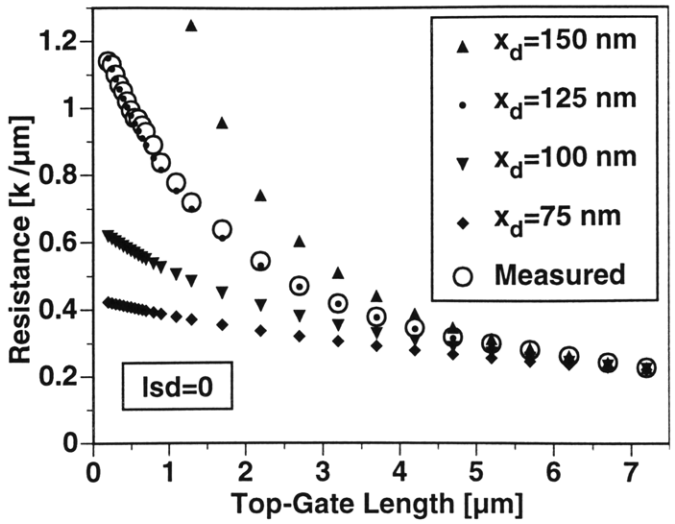


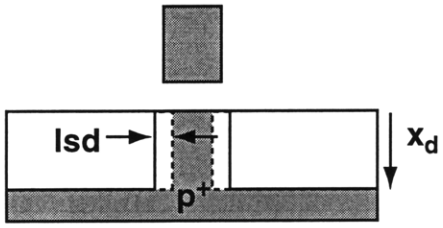
Figure 5-20. (a) Sample resistor *I-V* for a bottom-gate which has p-type counterdoped regions. (b) Curvature in the resistor *I-V* measurement is indicative of the n-type counterdoped regions.

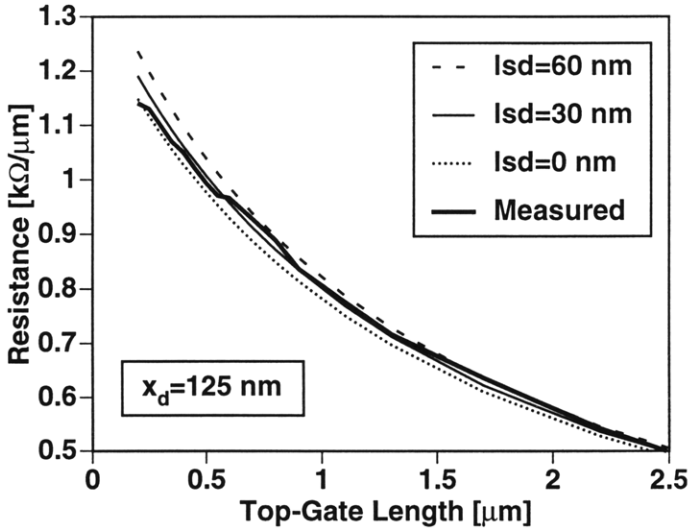
For various implant splits, the resistor *I-V*'s indicated whether the counterdope implant converted the counterdoped regions to a higher resistance p-type, or to n-type. This is shown schematically in Figure 5-20. Curvature in the resistor *I-V*'s indicate diode-like injection of holes into the floating of n-type regions. These holes are eventually recombined in the 100 μm-long n-type region. At low voltages however, the diode-action is not significant, so the low voltage portions of the *I-V* curves are used to extract the resistance which is to be used to characterize the effective





(a)





(b)

Figure 5-21. Using the low voltage resistance, the geometry of the counterdoped region can be modeled as a box with depth x_d. Lateral subdiffusion, *lsd*, does not affect long top-gate resistor devices so it is fixed at 0 nm when fitting x_d to measured data as shown in (a). Once the top-gate lengths get shorter, the curve fit is degraded indicating lateral subdiffusion is becoming important. (b) shows resistance versus top-gate length for shorter top-gate lengths for 3 values of lateral subdiffusion.

Bottom-Gate	B Dose [cm^{-2}]	As CD Dose [cm^{-2}]	B Conc. [cm^{-3}]	% B Activation	Curvature ?	x_d [nm]	lsd [nm]
recrystal	$3x10^{14}$	$1.5x10^{14}$	$4x10^{18}$	20	Y	125	30
recrystal	$3x10^{14}$	$3x10^{14}$	$4x10^{18}$	20	Y	122	60
recrystal	$1x10^{15}$	$5x10^{14}$	$1.5x10^{19}$	23	Y	125	30
recrystal	$1x10^{15}$	$1x10^{15}$	$1.5x10^{19}$	23	Y	115	90
polysilicon	$3x10^{14}$	$1.5x10^{14}$	$3x10^{17}$	1.5	N	110	30
polysilicon	$3x10^{14}$	$3x10^{14}$	$3x10^{17}$	1.5	Y	102	60
polysilicon	$1x10^{15}$	$5x10^{14}$	$5x10^{18}$	7.5	N	120	30
polysilicon	$1x10^{15}$	$1x10^{15}$	$5x10^{18}$	7.5	Y	115	40

Table 5-4. Summary of counterdoping experiment. Listed from left to right are the type of bottom-gate polysilicon, initial boron dose, arsenic counterdope dose, measured boron concentration, percent activation of boron, whether the resistor *I-V*'s show curvature, the extracted depth of counterdoping, and the extracted lateral subdiffusion length.

geometry of the counterdoped regions.

The model used is shown in Figure 5-21. The higher resistance counterdoped region is modeled as boxes of infinite resistance with depth x_d and width which is the difference between the bottom-gate and top-gate lengths. A lateral subdiffusion parameter, *lsd*, is added. For longer top-gate lengths, small *lsd* should not affect the fit of the model. Thus, the model is initially fitted by varying x_d while keeping *lsd* zero, as shown in Figure 5-21 (a). For short top-gate lengths, *lsd* becomes important, as shown in Figure 5-21 (b). The extracted x_d and *lsd* parameters are tabulated in Table 5-4, along with the original implant conditions, the polysilicon type, and whether the *I-V* curves had curvature indicative of n-type regions. Also calculated was the % activation of boron in the bottom-gate.

The resulting approximate as-implanted profiles show predictable trends. Boron activation is much higher in the recrystallized amorphous silicon than in fine-grain polysilicon and the counterdope implant depth is shallower in the fine-grain polysilicon than in the recrystallized amorphous silicon. Most of the samples showed evidence of n-type counterdoped regions, so the arsenic dose should be reduced. The *lsd* parameter is typically 30 nm. The higher numbers are for

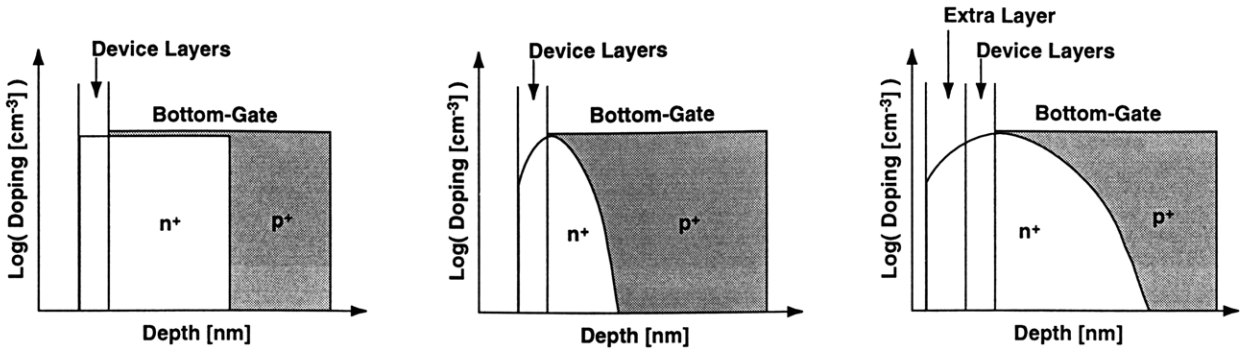


Figure 5-22. Doping concentration versus depth into the bottom-gate. (a) Ideal counterdoping implant profile is flat. (b) Ion implanted profiles are not flat. (c) Adding an extra layer into the implant can flatten out the ion implant profile.

higher doses which should not be used anyway given the counterdoped regions are already converted to n-type.

5.4.2 Discussion

Ideally, the implant should have zero lateral subdiffusion into the bottom-gate past the top-gate edge. It should be as flat as possible and as deep as possible, as shown in Figure 5-22 (a). In reality, the implant is sharply peaked as shown in Figure 5-22 (b). An extra layer, such as LTO, can be used to flatten out the implant profile, as shown in Figure 5-22 (c). This is still nowhere near ideal, so some sort of diffusion must be used to flatten out this profile. However, diffusion means there will be lateral subdiffusion. The approximate as-implanted profile characteristics extracted above already suggest that the devices are not scalable. With an approximate 30 nm lateral subdiffusion per side, a 60 nm top-gate length device would have little to no bottom-gate control.

Thus, any diffusion will not work unless spacers are used, as shown in Figure 5-23. This figure schematically compares the as-implanted counterdope depth and lateral profiles for a device without a spacer, with a spacer, and with an extra layer before spacer formation. In fact, this extra layer can be used to form the spacer itself. The spacer and the extra layer move the lateral profile away from the top-gate edge. The extra layer also flattens the depth profile. Thus, with either a

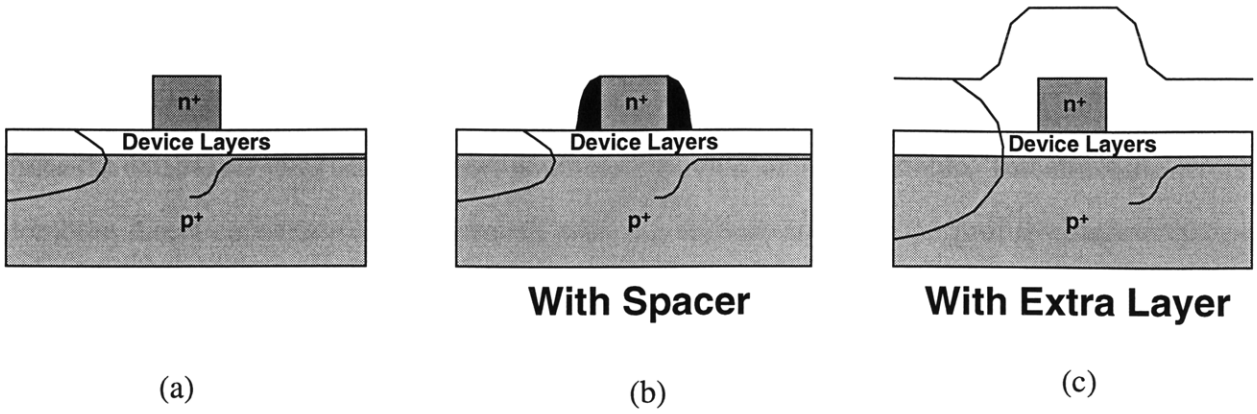


Figure 5-23. Schematic showing counterdoping species profiles. The vertical doping profile of the counterdoping species is shown on the left and the lateral doping profile is shown on the on the right for (a) just counterdoping through the device layers, (b) counterdoping with a spacer, and (c) counterdoping through an extra layer.

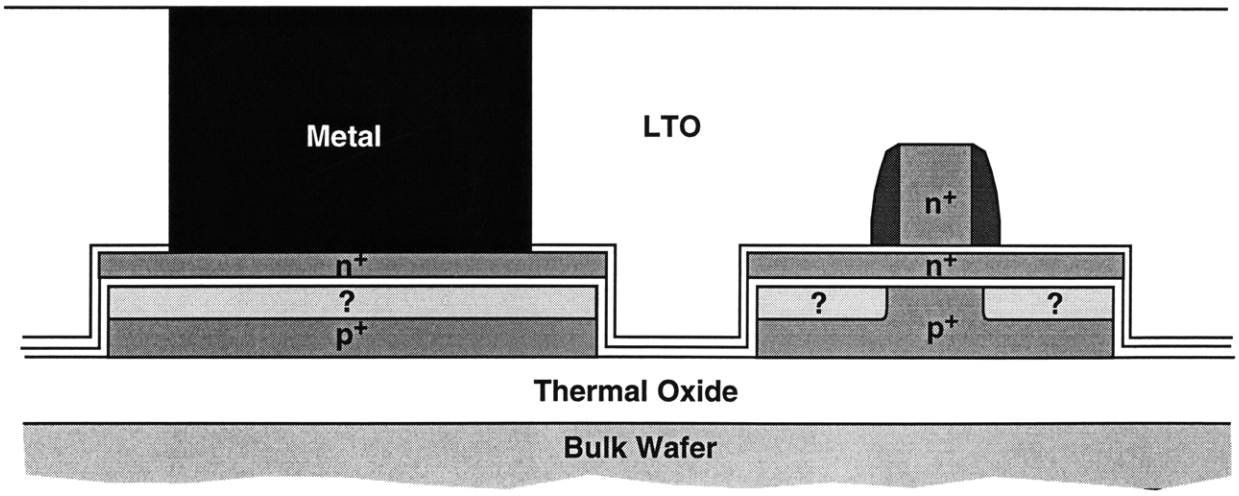


Figure 5-24. Proposed test structure cross section which can characterize the counterdoping profile with the resistor structure shown on the right, and characterize the effectiveness of counterdoping in reducing capacitance with the capacitor shown on the left.

spacer or an extra layer, a near ideal profile can be obtained which does not encroach laterally past the top-gate edge. However, much optimization is required, and this optimization will be rigorous since the extra layer thickness or spacer width become an additional variable. Test structures which combine direct capacitance measurement with the resistor counterdope profile characterization measurements described above should be used for such optimization. A cross section of possible test structure is illustrated in Figure 5-24.

5.5 Summary

A double-gate MOSFET fabrication technology was presented. The highlight of this fabrication technology was incorporation of a flip-bond-transfer process which buried patterned bottom-gates beneath an active silicon layer. NMOSFET's were fabricated as demonstration, and a method to self-align the top-gate to the bottom-gate by counterdoping the bottom-gate was explored.

Chapter 6

Conclusion

This thesis covered SOI MOSFET's scaled into the sub-100 nm technologies. It focused on partially-depleted SOI MOSFET design for sub-100 nm technologies. While many of the floating-body effects in SOI MOSFET's have been previously reported [11-19], much of this work was done on long channel devices and/or for the purpose of SPICE modeling. This thesis gives an in-depth treatment of device design based on understanding floating-body effects in the context of scaled CMOS, and in the context of digital logic. This treatment lead to the understanding of the "pulse stretching" phenomena, which was previously unknown. From this understanding, it was possible to establish design criteria for silicon film thickness. The silicon film thickness should be designed for "charge-balance" between devices starting in different logic states. With the silicon film thickness fixed by "charge-balance" for any combination of other design variables, it is possible to optimized PD-SOI MOSFET performance while minimizing floating-body hysteresis. And, with optimized PD-SOI MOSFET's, it was finally possible to fairly compare the intrinsic device performance of scaled bulk CMOS versus scaled SOI CMOS.

This thesis also presented a flip-bond-transfer process which was developed to fabricate double-gate SOI MOSFET's, which are essentially scaled fully-depleted SOI MOSFET's. This flip-bond-transfer process can be used for many other applications, as will be described below, particularly since the temperature of this process was pushed down to 600 °C. The following summarizes the key points of this thesis and discusses possible future applications of SOI technology.

6.1 Summary

The floating body voltage in partially-depleted SOI MOSFET's is capacitively coupled to transitions of the gate, source, drain, and substrate terminal voltages. The floating-body voltage changes instantaneously with changes in terminal voltage because the floating-body majority carrier content cannot change instantaneously, because there is no body contact. Over time however, the floating body majority carrier content will change from its equilibrium value due to impact ionization and reverse diode thermal generation, and forward bias diode and thermal recombination. Transient changes in body majority carrier content, and therefore body voltage, occurs until an equilibrium is established in which generation of body majority carriers is equal to recombination.

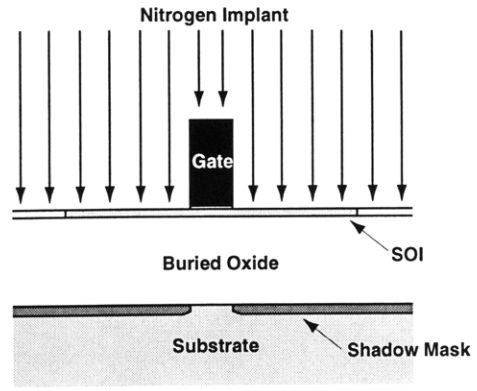
Because of the floating body, the instantaneous AC behavior of PD-SOI MOSFET's can be quite different from equilibrium DC behavior. This results in an equivalent model representing the PD-SOI MOSFET floating body which consists of body-source/drain diodes and impact ionization current which account for transient changes in body majority carrier content and body voltage, and body-terminal capacitances which account for instantaneous changes in body voltage. From this equivalent model, it was shown that only the instantaneous *I-V* characteristics are important for characterizing PD-SOI MOSFET switching behavior in digital circuits.

The instantaneous *I-V* behavior of PD-SOI MOSFET's are mainly characterized by the floating-body voltage in the OFF-state, and by capacitive coupling which boosts the body voltage during turn-ON of the MOSFET. Every time the MOSFET switches, the body majority carrier content changes, which changes the floating-body voltage in the OFF-state. While this is a minute

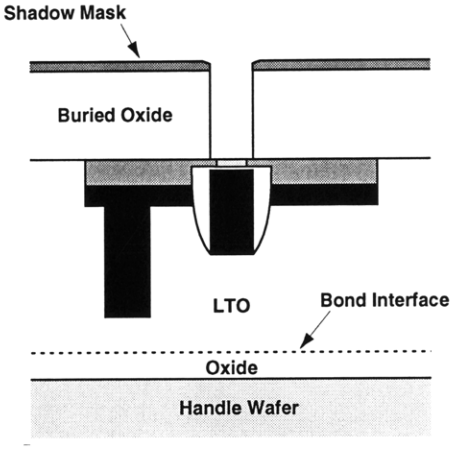
change per switch cycle, over many switching cycles the cumulative changes can be significant enough to noticeably alter instantaneous *I-V* behavior. This behavior of changing floating-body majority carrier content can be hysteretic, resulting in threshold voltage variation which becomes very important to circuit operation as power supplies are reduced -- it was shown that threshold voltage variation lead to an input pulse stretching or compressing as it propagated through a CMOS inverter chain, particularly at low power supply voltage. A simple guideline for designing the silicon film thickness for "charge-balance", that is, equal majority carrier content in equilibrium regardless of logic state was demonstrated to be effective in reduction of pulse stretching and minimizing hysteretic threshold voltage variation.

Once floating-body effects on *I-V* behavior and floating-body hysteresis were understood, it was possible to create an optimization framework for PD-SOI MOSFET design. It was shown that properly designed PD-SOI and bulk MOSFET's have comparable ON-currents and OFF-currents in sub-100 nm CMOS technologies. The ON-current at each gate length is degraded in SOI due to nonzero OFF-state floating-body voltage. However, this nonzero body bias reduces short-channel effects and allows SOI technologies to use shorter gate lengths as the nominal gate length. The shorter gate length SOI devices will have similar ON-current as bulk devices at a longer gate length, but the shorter gate length devices will have smaller gate capacitance. Thus, performance metrics for SOI technologies (such as CV_{DD}/I_{ON}) can be similar to or better than bulk technologies. In addition, PD-SOI still has the advantage over bulk with reduced parasitic source/drain capacitance and circuit-level advantages.

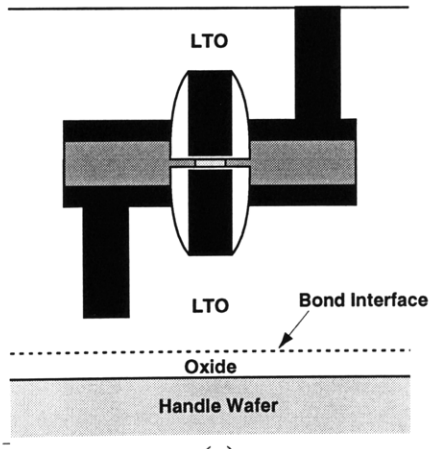
Fully-depleted SOI MOSFET's are expected to be even better than PD-SOI or bulk MOSFET's, provided a bottom-gate is in place to either maintain electrostatic integrity or act as the second gate in a double-gate MOSFET. A flip-bond-transfer process was demonstrated which can bury a prepatterned bottom-gate underneath a fully-depleted SOI MOSFET. The prepatterned bottom-gate process minimizes source/drain-bottom-gate capacitance. The prepatterned bottom-gate can be made self-aligned to the top-gate if it is doped the opposite type as the top-gate and then counterdoped through the top-gate. It was demonstrated, however, this counterdoping self-alignment



(a)



(b)



(c)

Figure 6-1. Self-aligned double-gate process proposed in [60]. (a) A chemically-modified buried layer is formed from the shadow of a gate. The top of the MOSFET is processed to metal, LTO is deposited and polished, and the wafer is flipped, bonded, and transfered onto a handle wafer. (b) The other gate is formed using the chemically-modified buried layer as a mask. (c) The second side of the MOSFET is processed.

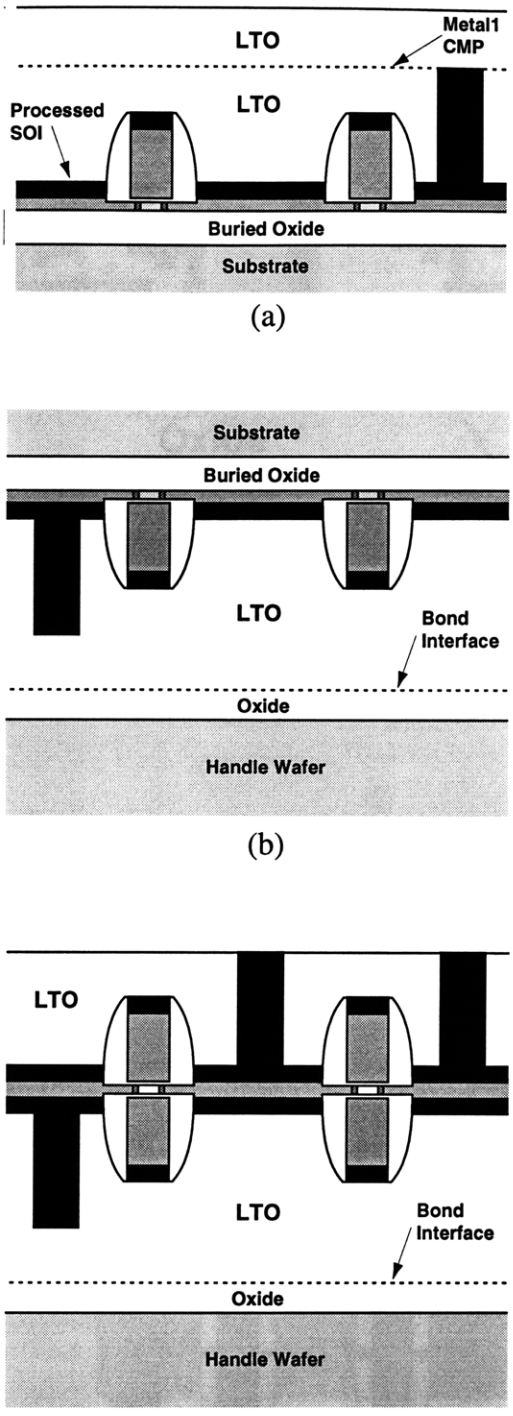


Figure 6-2. Schematic representation of layer transfer principles. This example is as follows: (a) one half of double-gate SOI devices been fabricated, a tungsten plug has been formed, and additional LTO has been deposited as a bonding layer. (b) the wafer is then flipped and bonded to a handle wafer. (c) the substrate and buried oxide are removed, and the other half of the double-gate SOI device is fabricated.

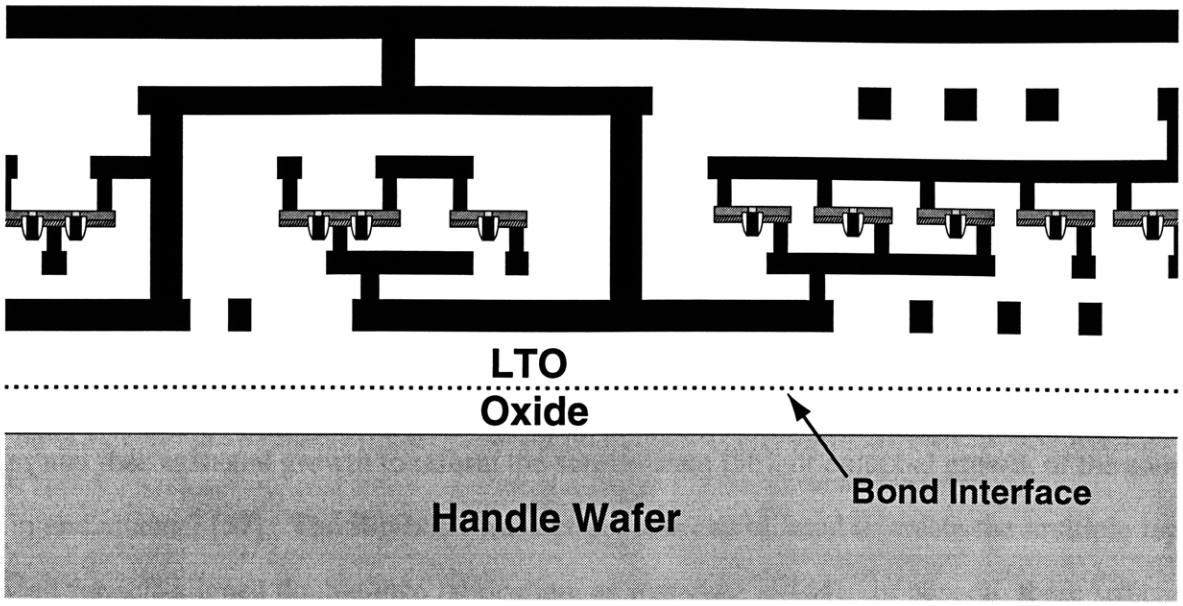


Figure 6-3. Schematic representation of double-sided interconnects fabricated with the flip-bond-delaminate process.

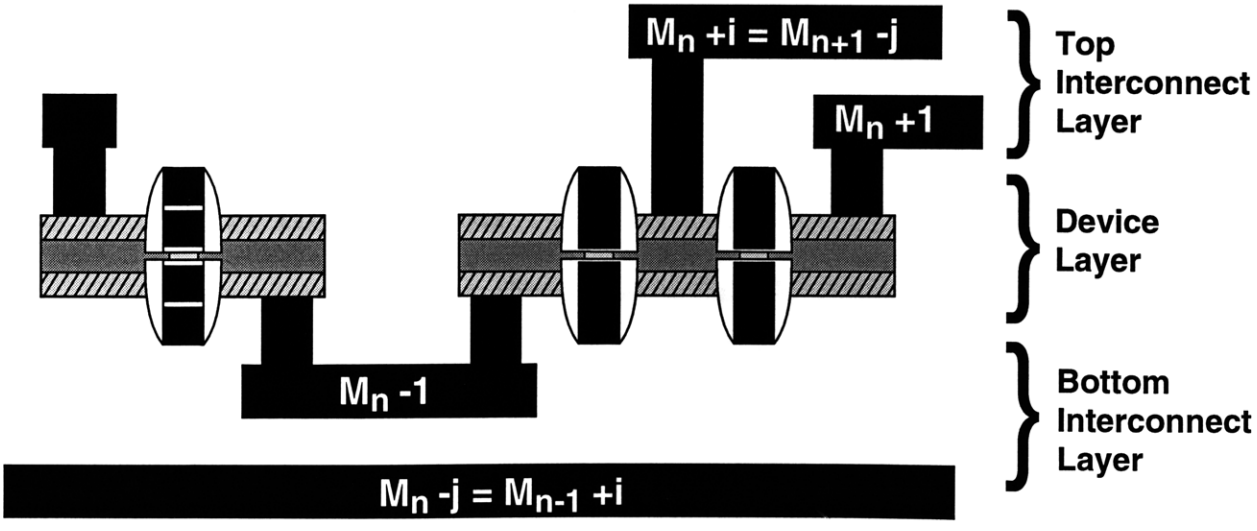


Figure 6-4. Schematic depiction of one "stratum" within a multi-strata 3-D integration. Shown is a device layer, i-levels of top interconnect, and j-levels of bottom interconnect.

scheme does not scale well since lateral encroachment of the counterdoping species due to diffusion and ion-implantation straggle does not scale.

6.2 Self-Aligned Double-Gate MOSFET's

A good topic for future work is fabrication of truly self-aligned double-gate MOSFET's. Several methods have been proposed. The first involve simultaneous etch of the top- and bottom-gates and then epitaxial growth to reform the source/drain [56], or epitaxial growth of the source/drain and channel [57]. The flip-bond-transfer process can be used to create the multiple layers needed for self-aligned double-gate fabrication, as was done in [44]. In general, these structures are quite difficult to fabricate.

The second type of process flips the MOSFET on its side producing either the FINFET [58], which is limited in gate width by the height of the structure, or a vertical MOSFET process, for which it is difficult to simultaneously build NMOS and PMOS [59]. Some process, which is inherently planar and which does not require re-forming parts of the MOSFET is preferable to those listed above. One such process, shown in Figure 6-1, shows some promise [60]. It involves implantation of some species which will form a masking layer after the flip-bond-delamination process. Details and variants of this process are still to be explored.

6.3 3-D SOI Integration

The flip-bond-transfer process shows promise for 3-D integration of SOI technologies. If a self-aligned double-gate process can incorporation metal layers into the flip-bond-transfer process as shown in Figure 6-2, then this process is the first step towards 3-D integration of 2-D planar processes. As shown in Figure 6-3, even if a self-aligned double-gate process is not used, the flip-bond-transfer process can be used to create double-sided interconnects. This will ease the wiring of transistors in future microprocessors. Ultimately, if layers can be stacked with flip-bond-transfer,

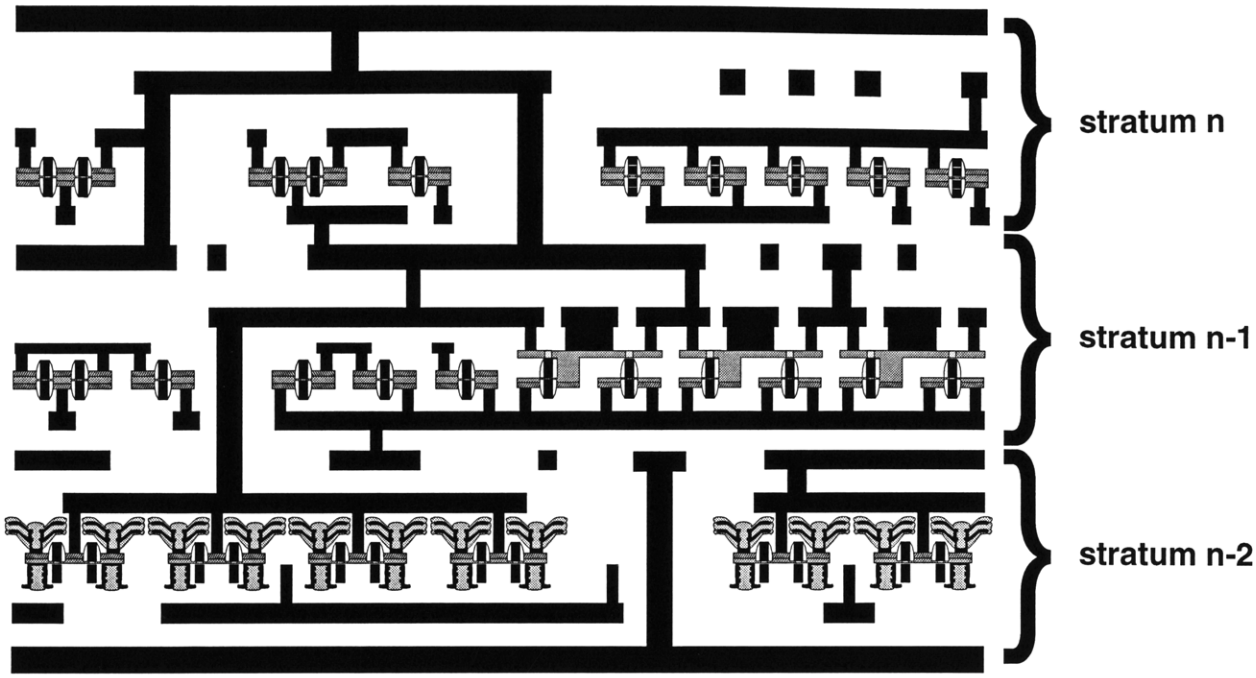


Figure 6-5. Schematic of 3-D "semiconductor in insulator" integration showing strata of device and interconnect layers.

or some other process, then we enter a new era of 3-D integration. A layer, or "stratum" can have multiple levels of interconnects as shown in Figure 6-4. Multiple strata of different technologies and materials can be integrated together, as shown in Figure 6-5. This integration of planar processes into 3-D systems will be better termed "semiconductor-in-insulator", and is ideal for system-on-a-chip applications.

Appendix A

Sample MEDICI Input Decks

The following are examples of input decks for the 2-D numerical simulator MEDICI. Descriptions of each file are shown in bold.

This input deck generates a simulation grid for 100 nm-technology bulk NMOSFET.

```
---------------------------------------------------------------------------
$ generates 100 nm technology bulk mesh with halo at xj
$ note Lgate is commented out because following file varies Lgate
$ filename=bulkgrid.mesh

$ doping paramters
assign name=SYsd n.val=1.954235e-02
assign name=Ah n.val=1.258574e+19
assign name=CYh n.val=3.250000e-02
assign name=SXh n.val=1.770361e-02
assign name=SYh n.val=2.577906e-02

assign  name=pdope   n.val=9e19
assign  name=bdope   n.val=1e15

$ physical dimensions
$$$$assign  name=Lgate       n.val=0.065
assign  name=Lud     n.val=0.0105
assign  name=Lsd     n.val=0.150-@Lgate/2

assign  name=Tpoly   n.val=0.090
```

```
assign  name=Tlto     n.val=0.025
assign  name=Tox      n.val=0.00155
assign  name=Tsi      n.val=0.200

$ tags
assign  name=xmin     n.val=-@Lgate/2-@Lsd
assign  name=xmax     n.val=@Lgate/2+@Lsd
assign  name=ymin     n.val=-@Tox-@Tpoly
assign  name=ymax     n.val=@Tsi

$ ********************************************************** $

mesh RECTANGU smooth.k=1

$ lateral mesh
x.mesh x.min=@xmin width=@Lsd h1=@Lsd/5 h2=@Lsd/25
x.mesh width=@Lgate/4 h1=@Lsd/25 h2=@Lgate/50
x.mesh width=@Lgate/4 h1=@Lgate/50 h2=@Lgate/25
x.mesh width=@Lgate/4 h1=@Lgate/25 h2=@Lgate/50
x.mesh width=@Lgate/4 h1=@Lgate/50 h2=@Lsd/25
x.mesh width=@Lsd h1=@Lsd/25 h2=@Lsd/5

$ depth mesh
y.mesh y.min=@ymin depth=@Tpoly h1=@Tpoly/3 h2=@Tox/3
y.mesh depth=@Tox h1=@Tox/2
y.mesh depth=@Tsi/4 h1=@Tox/4 h2=@Tsi/60
y.mesh y.max=@ymax h1=@Tsi/60 h2=@Tsi/10

eliminate       columns
+       x.min=@xmin x.max=-@Lgate/2 y.min=@ymin y.max=-@Tox
eliminate       columns
+       x.min=@Lgate/2 x.max=@xmax y.min=@ymin y.max=-@Tox

eliminate       columns
+       x.min=@xmin x.max=@xmax y.min=@ymax/2 y.max=@ymax

$ ********************************************************** $

$ left LTO and nitride
REGION      num=42 oxide
+       x.min=@xmin x.max=-@Lgate/2 y.min=-@Tox-@Tlto y.max=-@Tox
REGION      num=41 oxide
+       x.min=-@Lgate/2-@Tlto x.max=-@Lgate/2 y.min=@ymin y.max=-@Tox-@Tlto
REGION      num=43 nitride
+       x.min=@xmin x.max=-@Lgate/2-@Tlto y.min=@ymin y.max=-@Tox-@Tlto
```

```
$ right LTO and nitride
REGION      num=52 oxide
+       x.min=@Lgate/2 x.max=@xmax y.min=-@Tox-@Tlto y.max=-@Tox
REGION      num=51 oxide
+       x.min=@Lgate/2 x.max=@Lgate/2+@Tlto y.min=@ymin y.max=-@Tox-@Tlto
REGION      num=53 nitride
+       x.min=@Lgate/2+@Tlto x.max=@xmax y.min=@ymin y.max=-@Tox-@Tlto

$ gate polysilicon
region  num=1 silicon x.min=-@Lgate/2 x.max=@Lgate/2 y.min=@ymin y.max=-@Tox

$ gate oxide
region  num=2 oxide   x.min=@xmin x.max=@xmax y.min=-@Tox y.max=0

$ bulk silicon
region  num=3 silicon x.min=@xmin x.max=@xmax y.min=0 y.max=@ymax

$ ********************************************************** $

$ contacts
electrode name=drain x.min=@xmax-@Lsd/2 x.max=@xmax y.min=0 y.max=0

electrode name=gate  x.min=-@Lgate/2 x.max=@Lgate/2
+                    y.min=@ymin y.max=@ymin+@Tpoly/3

electrode name=source        x.min=@xmin x.max=@xmin+@Lsd/2 y.min=0 y.max=0

electrode name=bulk  bottom

$ ********************************************************** $

$ poly depletion
profile n-type region=1 uniform n.peak=@pdope

$ background doping
profile p-type region=3 uniform n.peak=@bdope

$ S/D 2D doping
profile n-type  n.peak=2.00e+20 region=3
+       y.min=0.00e-02 y.max=0.00e-02 y.char=@SYsd
+       x.min=@xmin x.max=-@Lgate/2 x.char=@Lud/1.5

profile n-type  n.peak=2.00e+20 region=3
+       y.min=0.00e-02 y.max=0.00e-02 y.char=@SYsd
```

```
+       x.min=@Lgate/2 x.max=@xmax x.char=@Lud/1.5

$ Halo 2D doping
profile p-type  n.peak=@Ah region=3
+       y.min=@CYh y.max=@CYh y.char=@SYh
+       x.min=@xmin x.max=-@Lgate/2+@Lud x.char=@SXh

profile p-type  n.peak=@Ah region=3
+       y.min=@CYh y.max=@CYh y.char=@SYh
+       x.min=@Lgate/2-@Lud x.max=@xmax x.char=@SXh

$ ********************************************************** $

contact name=gate     neutral resistan=0
contact name=bulk     neutral resistan=0

$ ********************************************************** $

extract mos.para
```

This input deck takes the grid generated above and generates gate transfer *I-V* characteristics for various gate lengths. The solution is based on hydrodynamic modeling.

```
---------------------------------------------------------------------
$ runs hydro idvgs sweep at vds=vdd for various gate lengths
$ this file calls bulkgrid.mesh

loop steps=8

assign  name=Lgate
+       n.val=(0.039,0.045,0.052,0.058,0.065,0.072,0.078,0.100)

assign  name=logfil c1=onoff039.log c2=onoff045.log c3=onoff052.log
+ c4=onoff058.log c5=onoff065.log c6=onoff072.log c7=onoff078.log
+ c8=onoff100.log

$ get grid
call file=bulkgrid.mesh

symb gummel carriers=0
solve init

$ models
models conmob unimob tmpmob ef.tmp auger
+ fermidir bgn
```

```
$ calibrated mobility for electrons, holes are defaults
mobility silicon betan=2.00 vsatn=1.00e7

$ energy relaxation time
material silicon ele.tauw=0.200e-12

symb  gummel carriers=1 electrons ele.temp
method  iccg damped itlimit=40
solve
symb  newton carriers=1 electrons ele.temp coup.ele
method  autonr n.damp
solve  v(drain)=0.01 vstep=0.01 nstep=3 elec=drain
solve  v(drain)=0.05 elec=drain
solve  v(drain)=0.06 vstep=0.01 nstep=3 elec=drain
solve  v(drain)=0.10 elec=drain
solve  v(drain)=0.125 vstep=0.025 nstep=2 elec=drain
solve  v(drain)=0.20 elec=drain
solve  v(drain)=0.25 elec=drain
solve  v(drain)=0.30 elec=drain
solve  v(drain)=0.35 elec=drain
solve  v(drain)=0.40 elec=drain
solve  v(drain)=0.45 elec=drain
solve  v(drain)=0.50 elec=drain
solve  v(drain)=0.55 elec=drain
solve  v(drain)=0.60 elec=drain
solve  v(drain)=0.65 elec=drain
solve  v(drain)=0.70 elec=drain
solve  v(drain)=0.75 elec=drain
solve  v(drain)=0.80 elec=drain
solve  v(drain)=0.85 elec=drain
solve  v(drain)=0.90 elec=drain
solve  v(drain)=0.95 elec=drain
solve  v(drain)=1.00 elec=drain
log  out.file=@logfil
solve
solve  v(gate)=0.01 vstep=0.01 nstep=3 elec=gate
solve  v(gate)=0.05 elec=gate
solve  v(gate)=0.06 vstep=0.01 nstep=3 elec=gate
solve  v(gate)=0.10 elec=gate
solve  v(gate)=0.125 vstep=0.025 nstep=2 elec=gate
solve  v(gate)=0.20 elec=gate
solve  v(gate)=0.25 elec=gate
solve  v(gate)=0.30 elec=gate
solve  v(gate)=0.35 elec=gate
solve  v(gate)=0.40 elec=gate
```

```
solve  v(gate)=0.45 elec=gate
solve  v(gate)=0.50 elec=gate
solve  v(gate)=0.55 elec=gate
solve  v(gate)=0.60 elec=gate
solve  v(gate)=0.65 elec=gate
solve  v(gate)=0.70 elec=gate
solve  v(gate)=0.75 elec=gate
solve  v(gate)=0.80 elec=gate
solve  v(gate)=0.85 elec=gate
solve  v(gate)=0.90 elec=gate
solve  v(gate)=0.95 elec=gate
solve  v(gate)=1.00 elec=gate

l.end

end
```

This input deck generates a grid to simulate a 100 nm-technology SOI NMOSFET.

```
---------------------------------------------------------------------------
$ generates 100 nm technology SOI mesh with halo at xj
$ filename=soigrid.mesh

assign  name=Lext     n.val=0.050
assign  name=Tsi      n.val=0.1200
assign  name=ptdope    n.val=6e17
assign  name=pttop     n.val=0.0600

assign  name=Lgate    n.val=0.065
assign name=Lscale n.value=@Lgate/0.005

assign name=SYsd n.val=1.954235e-02
assign name=Ah n.val=1.258574e+19
assign name=CYh n.val=3.250000e-02
assign name=SXh n.val=1.770361e-02
assign name=SYh n.val=2.577906e-02

assign  name=bdope   n.val=1e15
assign  name=pdope   n.val=9e19

$ physical dimensions
assign  name=Ldsd     n.val=0.050

$ Lud is Lgate-Lmet, it is the underdiffusion, the offset between the halo
$ edge and the extension edge
assign  name=Lud     n.val=0.0105
```

```
assign  name=Tpoly   n.val=0.090
assign  name=Tlto    n.val=0.025
assign  name=Tox     n.val=0.00155
assign  name=Tbox      n.val=0.1000

$ tags
assign  name=xmin    n.val=-@Lgate/2-@Lext-@Ldsd
assign  name=xmax    n.val=@Lgate/2+@Lext+@Ldsd
assign  name=ymin    n.val=-@Tox-@Tpoly
assign  name=ymax    n.val=@Tsi

$ ********************************************************** $

mesh RECTANGU smooth.k=1 ^diag.flip

$$$$$ for Lext > 0 nm $$$$$$
if cond=(@Lext>0)
x.mesh  x.min=@xmin width=@Ldsd/2 h1=0.01 h2=0.01
x.mesh  width=@Ldsd/2+@Lext/2 h1=0.01 h2=0.005
x.mesh  width=@Lext/2 h1=0.005 h2=@Lgate/@Lscale
x.mesh  width=@Lgate/2 h1=@Lgate/@Lscale h2=@Lgate/@Lscale
x.mesh  width=@Lgate/2 h1=@Lgate/@Lscale h2=@Lgate/@Lscale
x.mesh  width=@Lext/2 h1=@Lgate/@Lscale h2=0.005
x.mesh  width=@Ldsd/2+@Lext/2 h1=0.005 h2=0.01
x.mesh  width=@Ldsd/2 h1=0.01 h2=0.01
if.end

$$$$$ for Lext = 0 nm $$$$$$
if cond=(@Lext=0)
x.mesh  x.min=@xmin width=@Ldsd/2 h1=0.01 h2=0.01
x.mesh  width=@Ldsd/2 h1=0.01 h2=0.005
x.mesh  width=@Lgate/2 h1=@Lgate/@Lscale h2=@Lgate/@Lscale
x.mesh  width=@Lgate/2 h1=@Lgate/@Lscale h2=@Lgate/@Lscale
x.mesh  width=@Ldsd/2 h1=0.005 h2=0.01
x.mesh  width=@Ldsd/2 h1=0.01 h2=0.01
if.end

$$$$$ for Tsi < 500 A $$$$$$
if cond=(@Tsi<0.0500)
y.mesh  y.min=@ymin depth=@Tpoly h1=@Tpoly/3 h2=@Tox/3
y.mesh  depth=@Tox h1=@Tox/2
y.mesh  depth=@Tsi h1=0.002 h2=0.002
y.mesh  depth=@Tbox h1=0.01 h2=0.05
if.end
```

```
$$$$$ for 800 A > Tsi >= 500 A $$$$$$
if cond=(@Tsi<0.0800)&(@Tsi>=0.0500)
y.mesh  y.min=@ymin depth=@Tpoly h1=@Tpoly/3 h2=@Tox/3
y.mesh  depth=@Tox h1=@Tox/2
y.mesh  depth=0.0350 h1=0.002 h2=0.003
y.mesh  depth=@Tsi-0.0350 h1=0.003 h2=0.002
y.mesh  depth=@Tbox h1=0.01 h2=0.05
if.end

$$$$$ for Tsi >= 800 A $$$$$$
if cond=(@Tsi>=0.0800)
y.mesh  y.min=@ymin depth=@Tpoly h1=@Tpoly/3 h2=@Tox/3
y.mesh  depth=@Tox h1=@Tox/2
y.mesh  depth=0.0350 h1=0.002 h2=0.003
y.mesh  depth=@Tsi-0.0700 h1=0.003 h2=0.003
y.mesh  depth=0.0350 h1=0.003 h2=0.002
y.mesh  depth=@Tbox h1=0.01 h2=0.05
if.end

$ ********************************************************** $

$ gate polysilicon
region  num=1 silicon x.min=-@Lgate/2 x.max=@Lgate/2 y.min=@ymin y.max=-@Tox

$ gate oxide
region  num=2 oxide x.min=@xmin x.max=@xmax y.min=-@Tox y.max=0

$ SOI silicon
region  num=3 silicon x.min=@xmin x.max=@xmax y.min=0 y.max=@ymax

$ buried oxide
region  num=4 oxide x.min=@xmin x.max=@xmax y.min=@Tsi

$ left LTO
REGION      num=5 oxide
+       x.min=@xmin x.max=-@Lgate/2 y.min=@ymin y.max=-@Tox

$ right LTO and nitride
REGION      num=6 oxide
+       x.min=@Lgate/2 x.max=@xmax y.min=@ymin y.max=-@Tox

$ ********************************************************** $

$ contacts
```

```
electrode name=gate x.min=-@Lgate/2 x.max=@Lgate/2
+               y.min=@ymin y.max=@ymin+@Tpoly/3
electrode name=source x.min=@xmin x.max=@xmin y.min=0 y.max=@Tsi
electrode name=drain x.min=@xmax x.max=@xmax y.min=0 y.max=@Tsi
electrode name=substrate bottom

$ body contact
$electrode name=body x.min=-@Lgate/8 x.max=@Lgate/8 y.min=@Tsi y.max=@Tsi

$ ********************************************************** $

$ poly depletion
profile n-type region=1 uniform n.peak=@pdope

$ background doping
profile p-type region=3 uniform n.peak=@bdope

$ punchthrough doping
if cond=(@Tsi>@pttop)
profile p-type region=3 n.peak=@ptdope x.char=@SXh y.char=@SYh
+ x.min=@xmin x.max=@xmax y.min=@pttop y.max=@ymax
if.end

$ Left Deep S/D, Extension S/D, and Halo
profile n-type  n.peak=2.000000e+20 region=3
+     y.min=0.000000e+00 y.max=@Tsi y.char=@Lud/1.5
+     x.min=@xmin x.max=@xmin+@Ldsd x.char=@Lud/1.5

profile n-type  n.peak=2.000000e+20 region=3
+     y.min=0.000000e+00 y.max=0.000000e+00 y.char=@SYsd
+     x.min=@xmin x.max=-@Lgate/2 x.char=@Lud/1.5

profile p-type  n.peak=@Ah region=3
+     y.min=@CYh y.max=@CYh y.char=@SYh
+     x.min=@xmin x.max=-@Lgate/2+@Lud x.char=@SXh


$ Right Halo, Extension S/D, Deep S/D
profile p-type  n.peak=@Ah region=3
+     y.min=@CYh y.max=@CYh y.char=@SYh
+     x.min=@Lgate/2-@Lud x.max=@xmax x.char=@SXh

profile n-type  n.peak=2.000000e+20 region=3
+     y.min=0.000000e+00 y.max=0.000000e+00 y.char=@SYsd
+     x.min=@Lgate/2 x.max=@xmax x.char=@Lud/1.5
```

```
profile n-type  n.peak=2.000000e+20 region=3
+       y.min=0.000000e+00 y.max=@Tsi y.char=@Lud/1.5
+       x.min=@xmax-@Ldsd x.max=@xmax x.char=@Lud/1.5

$ ************************************************************ $

contact name=gate  neutral resistan=0
contact name=substrate  neutral resistan=0

$ ************************************************************ $

regrid  reg=3 doping ratio=2 log smooth.k=1

extract mos.para
```

This input deck takes the SOI grid generated above, steps the drain voltage to V_{DD}, then rapidly ramps the gate voltage. Solutions are saved at each timestep. The body voltage can be extracted from each solution.

```
---------------------------------------------------------------------
$ this file calculates transient body voltage boost

$ get grid
call file=soigrid.mesh
save mesh out.file=mesh.mesh

extract mos.para

$ models
models fermidir conmob unimob fldmob consrh bgn auger r.tunnel

$ calibrated mobility for electrons, holes are defaults
mobility silicon betan=2.00 vsatn=1.00e7

symb gummel carriers=0
solve init
symb  newton carriers=2
method  autonr n.damp
solve  v(drain)=0.01 vstep=0.01 nstep=3 elec=drain
solve  v(drain)=0.05 elec=drain
solve  v(drain)=0.06 vstep=0.01 nstep=3 elec=drain
solve  v(drain)=0.10 elec=drain
solve  v(drain)=0.125 vstep=0.025 nstep=2 elec=drain
solve  v(drain)=0.20 elec=drain
solve  v(drain)=0.25 elec=drain
```

```
solve  v(drain)=0.30 elec=drain
solve  v(drain)=0.35 elec=drain
solve  v(drain)=0.40 elec=drain
solve  v(drain)=0.45 elec=drain
solve  v(drain)=0.50 elec=drain
solve  v(drain)=0.55 elec=drain
solve  v(drain)=0.60 elec=drain
solve  v(drain)=0.65 elec=drain
solve  v(drain)=0.70 elec=drain
solve  v(drain)=0.75 elec=drain
solve  v(drain)=0.80 elec=drain
solve  v(drain)=0.85 elec=drain
solve  v(drain)=0.90 elec=drain
solve  v(drain)=0.95 elec=drain
solve  v(drain)=1.00 elec=drain

solve  v(gate)=1.0 elec=gate ramptime=50e-12 dt=1e-14 tstop=100e-9
+ out.file=soln.aa

end
```

This input deck extracts the body voltage from each solution generated above.

```
---------------------------------------------------------------------
$ this file extracts the body voltage from the transient solutions

$****** Silicon Film Thickness *************
assign  name=Tsi n.value=0.1200
$*****************************************

mesh in.file=mesh.mesh
loop step=100
assign name=solnfile c.value=soln.aa delta=1
load in.file=@solnfile
print solution x.min=0.0 x.max=0.0 y.min=@Tsi y.max=@Tsi
l.end

end
```

This input deck generates the DC solution starting OUT-HI. It takes the SOI grid above and steps the drain voltage to V_{DD} while the gate voltage is 0 V.

```
---------------------------------------------------------------------
$ this file generates initial DC solution starting OUT-HI

$ get grid
call file=soigrid.mesh
```

```
save mesh out.file=mesh.mesh

extract mos.para

$ models
models fermidir conmob unimob fldmob consrh bgn auger r.tunnel

$ calibrated mobility for electrons, holes are defaults
mobility silicon betan=2.00 vsatn=1.00e7

symb gummel carriers=0
solve init
symb  newton carriers=2
method  autonr n.damp
solve  v(drain)=0.01 vstep=0.01 nstep=3 elec=drain
solve  v(drain)=0.05 elec=drain
solve  v(drain)=0.06 vstep=0.01 nstep=3 elec=drain
solve  v(drain)=0.10 elec=drain
solve  v(drain)=0.125 vstep=0.025 nstep=2 elec=drain
solve  v(drain)=0.20 elec=drain
solve  v(drain)=0.25 elec=drain
solve  v(drain)=0.30 elec=drain
solve  v(drain)=0.35 elec=drain
solve  v(drain)=0.40 elec=drain
solve  v(drain)=0.45 elec=drain
solve  v(drain)=0.50 elec=drain
solve  v(drain)=0.55 elec=drain
solve  v(drain)=0.60 elec=drain
solve  v(drain)=0.65 elec=drain
solve  v(drain)=0.70 elec=drain
solve  v(drain)=0.75 elec=drain
solve  v(drain)=0.80 elec=drain
solve  v(drain)=0.85 elec=drain
solve  v(drain)=0.90 elec=drain
solve  v(drain)=0.95 elec=drain
solve  v(drain)=1.00 elec=drain
save solution out.file=outhi.soln

end
```

This input deck runs a transient *I-V* sweep using the OUT-HI solution generated above. A transient *I-V* sweep is done by ramping the gate voltage. The transient drain current is picked off the logfile a few timesteps after the end of the ramp, after the displacement current has disappeared.

```
$ this file runs transient I-V for starting OUT-HI
title  does OUT-HI (VDD=1.0) trans-iv sweep

mesh  in.file=mesh.mesh

loop  steps=7

assign  name=vg n.value=(0.01,0.1,0.2,0.3,0.4,0.5,0.6)
assign  name=logfil c1=outhivg001.log c2=outhivg010.log c3=outhivg020.log
+ c4=outhivg030.log c5=outhivg040.log c6=outhivg050.log c7=outhivg060.log
+ c8=outhivg070.log c9=outhivg150.log

load  in.file=outhi.soln
models fermidir conmob unimob fldmob consrh bgn auger r.tunnel
symb  newton carriers=2
log  out.file=@logfil
solve  v(gate)=@vg elec=gate ramptime=50e-12 dt=1e-14 tstop=100e-9

l.end

end
```

This input deck takes the SOI grid above and generates a OUT-HI solution, but starting from a OUT-LO initial DC solution. The gate is stepped to V_{DD} with the drain at 0 V. The gate voltage is then quickly ramped to 0 V and the drain is quickly ramped to V_{DD}. This results in negligible change in body hole content.

```
---------------------------------------------------------------------
$ this file generates initial DC solution starting OUT-LO, then flips to OUT-HI

$ get grid
call file=soigrid.mesh
save mesh out.file=mesh.mesh

extract mos.para

$ models
models fermidir conmob unimob fldmob consrh bgn auger r.tunnel

$ calibrated mobility for electrons, holes are defaults
mobility silicon betan=2.00 vsatn=1.00e7

symb gummel carriers=0
solve init
solve  v(gate)=0.01 vstep=0.01 nstep=3 elec=gate
solve  v(gate)=0.05 elec=gate
```

```
solve  v(gate)=0.06 vstep=0.01 nstep=3 elec=gate
solve  v(gate)=0.10 elec=gate
solve  v(gate)=0.125 vstep=0.025 nstep=2 elec=gate
solve  v(gate)=0.20 elec=gate
solve  v(gate)=0.25 elec=gate
solve  v(gate)=0.30 elec=gate
solve  v(gate)=0.35 elec=gate
solve  v(gate)=0.40 elec=gate
solve  v(gate)=0.45 elec=gate
solve  v(gate)=0.50 elec=gate
solve  v(gate)=0.55 elec=gate
solve  v(gate)=0.60 elec=gate
solve  v(gate)=0.65 elec=gate
solve  v(gate)=0.70 elec=gate
solve  v(gate)=0.75 elec=gate
solve  v(gate)=0.80 elec=gate
solve  v(gate)=0.85 elec=gate
solve  v(gate)=0.90 elec=gate
solve  v(gate)=0.95 elec=gate
solve  v(gate)=1.00 elec=gate

symb  newton carriers=2
method  autonr n.damp
solve  v(drain)=1.0 elec=gate ramptime=10e-12 dt=1e-14 tstop=20e-12
solve  v(gate)=0 elec=gate ramptime=10e-12 dt=1e-14 tstop=40e-12
save solution out.file=outlo.soln

end
```

This input deck runs a transient *I-V* on the solution starting OUT-LO, as generated above.

```
---------------------------------------------------------------------
$ this file runs transient I-V for starting OUT-LO, right after flipping OUT-HI
title  does OUT-LO (VDD=1.0) trans-iv sweep

mesh  in.file=mesh.mesh

loop  steps=7

assign  name=vg n.value=(0.01,0.1,0.2,0.3,0.4,0.5,0.6)
assign  name=logfil c1=outlovg001.log c2=outlovg010.log c3=outlovg020.log
+ c4=outlovg030.log c5=outlovg040.log c6=outlovg050.log c7=outlovg060.log
+ c8=outlovg070.log c9=outlovg150.log

load  in.file=outlo.soln
models fermidir conmob unimob fldmob consrh bgn auger r.tunnel
```

```
symb  newton carriers=2
log  out.file=@logfil
solve  v(gate)=@vg elec=gate ramptime=50e-12 dt=1e-14 tstop=100e-9

l.end

end
```

This is a sample MEDICI Circuit Advanced Application Module (CAAM) file in which the gate and drain voltage change simulatneously. This is done by defining the gate and drain voltages as SPICE elements connected to the simulation mesh. The body hole content can be increased if the simulation mesh has a floating body by using the "photogen" statement. A SPICE current source can be added to the body terminal if the simulation mesh is defined with a body contact. This file generates solutions versus time. Body voltage, hole content, etc., can be extracted from these solutions.

```
$ this is a sample CAAM file in which gate and drain voltages are varied
$ simultaneously to mimic switching in a CMOS inverter
$ C-AAM Deck for PD SOI CMOS Inverter
$assign  name=tsi n.value=
$assign  name=tox n.value=
$assign  name=leff n.value=
$assign  name=na n.value=
$assign  name=gfac n.value=
call    file=channel

start  circuit

$ nodes: 0-ground
$       1-Vin (gate)
$       2-Vout (drain)

$ pulse:      | v0| val  td|  tr|  tf|  tp| per|
vin  1 0 pulse  0   1.2    0  50p  50p   1n  2n
vd   2 0 pulse  1.2   0  30p  50p  50p   1n  2n

$ node=electrode name
pnmos  1=gate 0=source 2=drain 0=substrate file=mesh.mesh width=1.0

$ set up initial conditions on the nodes
.nodeset  v(1)=0 v(2)=1.2

$ set maximum voltage step to be less than default
.option delvmax=0.2
```

```
finish  circuit

material  print
save  mesh out.file=caam.mesh

load  in.file=outhi120.transsoln structure=pnmos
extract holes structure=pnmos
+ x.min=0.235 x.max=0.235+@leff y.min=0.0 y.max=@tsi/1e4

models  arora hpmob consrh auger bgn impact.i
symbol  newton carr=2
log  ivfile=log.log
solve  dt=1e-14 tstop=1.2e-9 out.file=soln.aa impact.i

end
```

The following is a section of an input deck which extracts holes content from solutions generated by MEDICI CAAM.

```
---------------------------------------------------------------------
$ this is a sample CAAM file in which extract holes from CAAM solutions
$assign  name=tsi n.value=
$assign  name=tox n.value=
$assign  name=leff n.value=
$assign  name=na n.value=
$assign  name=gfac n.value=
call    file=channel

start circuit
.load mesh=caam.mesh solution=soln.aa
finish circuit
extract structure=pnmos holes
+ x.min=0.235 x.max=0.235+@leff y.min=0.0 y.max=@tsi/1e4

start circuit
.load mesh=caam.mesh solution=soln.ab
finish circuit
extract structure=pnmos holes
+ x.min=0.235 x.max=0.235+@leff y.min=0.0 y.max=@tsi/1e4

etc.
```

Appendix B

Double-Gate MOSFET Process Traveler

Below is the detailed process traveler for the double-gate MOSFET process described in section 5.1. Each enumerated step is divided into detailed substeps, each with the appropriate MTL machine name in parentheses. This traveler is presented as reference only. Many of the machines listed and many of the corresponding recipes may no longer be in use due to conversion of the fabrication facilities from 4-inch equipment to 6-inch equipment.

Lot Name: NVTMOS4

Description: Self-aligned double-gate NMOS process with prepatterned backgate and self-alignment through counterdoping of the backgate.

Maskset: COMPBG1
CMP-CPB(clear)-backpoly
CD(clear)-active
CP(clear)-poly
CCB(dark)-back contact cuts
CC(dark)-contact cuts
CI(clear)-implant mask (blocks n+)
CI-comp(dark)-implant mask (blocks p+)
CNC(clear)-implant mask (blocks n+ counterdoping implant)
CM(clear)-metal mask

****

1. Thin SOI Wafers
****
Starting 1950A, we want to finish with 400A. Assume 0.5 consumption to be safe; also, SOI oxidizes on the back interface a little too. Oxidation steps in the process include backgateoxide(200A), SRO(250A), sacox(60A), frontgateoxide(50A), reox(~30A on top of gateox that's already there), so the total consumed is 590A*0.5=295A.
1950-400-295=1255A/0.5=2510A oxide grown to thin wafers.
a) rca add 2 blanks (rca)
---
b) grow 0.5*2510A oxide, add 2 blanks (tubeB1 rec. 139, FF=2hr07min)
***load in tube in order, unload in order
---
c) BOE dip to remove oxide (oxide)
MACHINE: ICL oxide DATE: 02/26/99 10:34:22 USER: wei
---
d) rca add 2 blanks (rca)
MACHINE: ICL rca DATE: 02/26/99 11:50:20 USER: wei
---
e) grow 0.5*2510A oxide (tubeB1 rec. 139, FF=2hr01min)
***load in tube in opposite order to even out positional variation in boat
---
f) BOE dip to remove oxide (oxide)

****
2. Back-Gate Stack and Make Handle Wafers
****
grow back-gate oxide on SIMOX wafers and deposit back-poly
a) rca add 2 blanks (rca) device wafers and handle wafers
---
b) grow ~5k wet thermal oxide on handle wafers (tubeB1, rec. 122 FF=1hr35min)
---
c) back-gate oxidation 200A 900C dry oxide (tubeA2 rec. 226 FF=1hr16min)
---
d) deposit 1750A a-Si (tubeA6 rec. 705 FF=1hr50min) add 4 monitors

****
3. Implant Back-Gate
****
a) blanket implant back gate poly p+ dose=1.5e15 (Ion Implant Services)
---
b) piranah clean (pre-metal strip)

****
4. Back-Gate Definition

****
stepper2 pattern CMP-CPB job WEI CPB 1st level, shoot 9x9
---
b) coat (coater)
---
c) shoot 9x9 (stepper2)
---
d) develop (developer)
---
e) BOE dip before poly etch (oxide)
---
f) etch back-poly (AME5000 chamberB recipe UNDOPED POLY NOD)
---
g) dump rinse in undoped oxide dump rinser to remove chlorine (oxide)
---
h) ash resist (asher)

****
5. Prepare Wafers for Bonding
****
a) rca NO HF DIP (rca)
---
b) deposit 1.5um LTO (tubeA7, rec. 462 FF=2hr25min)
---
c) poly recrystallization anneal (tubeB4 rec. 176 600C FF=14hr)
---
d) densify LTO 1000C (tubeA3, rec. 322, FF=10min)
---
e) CMP SOI wafers (CMP)
---
f) post-CMP clean (pre-metal strip)

****
6. Bond Wafers
****
a) rca clean (rca)
---
b) SC-1 clean and bond (rca and bonding jig)
---
c) anneal bond 950C (tubeA3, rec. 208, FF=40min)
---
d) deposit ~1100A nitride (tubeA5, rec. 460 FF=43min)
---
e) deposit ~1100A nitride (tubeA5, rec. 460 FF=43min), rotate wafers 180 degrees so the parts in the rail are not exposed

****
7. Etchback Bonded Wafers
****
a) thin wafers (grind and polish, Skip Hoyt, Lincoln Labs, 181-4433)
---
b) piranah and 50:1 HF dip in yellow flourowre in TRL, then piranah
in green flouroware (acid-hood)
---
c) piranah clean (pre-metal clean)
---
d) deposit 5um 25% lf pwr TEOS on backside (concept1)
---
e) coat back handle wafer (coater) program 33 no EBR
---
f) BOE dip off TEOS that spilled over to TMAH side (oxide)
---
g) ash resist (asher)
---
h) TMAH wet etch of SIMOX substrate (KOHhood)
---
i) post-TMAH piranah with HF dip (TRL acid-hood yellow flouroware, then green)
this is to clean wafers and knock loose particles off of void edges
---
j) coat frontside (coater) program 11, want EBR to remove TEOS off edges
---
k) BOE dip to strip TEOS off backside (oxide)
---
l) scribe wafers (scribe in TRL)
---
m) piranah strip resist (pre-metal)
---
n) BOE dip SOI wafers to remove remaining SIMOX BOX (oxide)
---
o) rca clean alignment dummies (rca)
---
p) deposit 1k a-Si (tubeA6 rec. 705 FF=59min, based on last run rate=17A/min)
---
q) inspect

****
8. SRO/Nitride
****
a) rca clean (rca)
---

b) grow 220A SRO (tubeA1 rec. 230) add 2 blanks and 2 blanks as etch dummies
---
c) deposit 1450A nitride (tubeA5 rec. 410) add 4 blanks as etch dummies and
1 1000A monitor in case we have to stip nitride

****
9. Define Active Area
****
stepper2 pattern CD, align to back poly
a) HMDS (HMDS)
---
b) coat (coater)
---
c) shoot (stepper2) job WEI FILL,1
---
d) develop (developer)
---
e) nitride plasma etch (AME5000 chamberA, recipe NITRIDE ETCH CF4)
---
f) field implant (Ion Implant Services)
5E13 Boron 10keV tilt=7 on machine 6001 for double-gate wafers
5E13 Boron 25keV tilt=7 on standard CF machine for bulk wafers
---
g) piranah clean (pre-metal strip)
---
h) ash resist (asher)
resist stripped in piranah

****
10. Field Oxide
****
a) rca clean (rca)
---
b) field oxidation (tubeB1 rec. 114 FF=27min for SOI, 1hr35min for bulk)
tsi based on 50% consumption thus far is 485A, thus FOX needed based on 44%
consumption+20% is 1323A, aim high at 1350A, FOX for bulk wafers will be 3400A,
based on dummy run SOI gets FF=27min in tubeB1 and bulk gets FF=1hr35min in
tubeB2

****
11. Strip Nitride/SRO and Grow SacOx
****
a) 5sec BOE dip before nitride wet etch (oxide) minimize dip to miminize
loss of FOX
---

b) nitride wet etch (nitride)
---
c) BOE dip off SRO (oxide) minimize dip to miminize loss of FOX
---
d) rca clean (rca)
---
e) grow ~60A dummy gate oxide at 900C (tubeA2 rec. 226 FF=8min)

****
12. Vt Adjust Implants (for short channel Vt=0.1V)
****
a) implant channel BF2 1.5E12 25keV tilt=7 (Ion Implant Services)
bulk wafers also got Boron 5E12 50keV tilt=7
---
b) piranah clean (pre-metal strip)
---
c) BOE dip to remove dummy gate oxide (oxide)

****
13. Form Gate Stack
****
a) rca clean (rca)
---
b) grow 60A gate oxide at 900C (tubeA2 rec. 226 FF=7min30sec)
---
c) depsoit as polysilicon (tubeA6 rec. 461 FF=30min)

****
14. Poly Definition
****
stepper2 pattern CP job WEI FILL align to backgate
---
b) coat (coater)
---
c) shoot (stepper2)
---
d) develop (developer)

****
15. Etch Poly
****
a) BOE dip 5sec to clear native oxide (oxide)
---
b) poly plasma etch (AME5000 chamberB recipe UNDOPED POLY NOD)
---

c) dump rinse after chlorine etch (oxide)
---
d) ash resist (asher)

****
16. Reox
****
a) rca NO HF DIP (rca) add 2 blanks
---
b) grow 900C 60A oxide (tubeB2 rec. 226 FF=6min)

****
17. Back Contact Cuts
****
stepper2 pattern CCB, align to back poly
---
b) coat (coater)
---
c) shoot (stepper2)
---
d) develop (developer)
---
e) BOE dip to open back contacts (oxide)
---
f) back contact implant Boron 3E15 10keV 7 degree tilt (Ion Implant Services)
---
g) piranah clean (pre-metal strip)
---
i) ash resist (asher)

****
18. Source/Drain/Top-Gate Implants
****
stepper2 pattern CI, align to back poly, shoot SOI wafers only
a) HMDS (HMDS)
---
b) coat (coater)
---
c) shoot (stepper2)
---
d) develop (developer)
---
e) implant top poly and S/D (Ion Implant Services) send out SOI and bulk wafers
Arsenic, 4E15, 25keV, 0 degree tilt
---

f) piranah clean (pre-metal strip)

---

g) ash resist (asher)

****

19. Source/Drain Top-Gate and Bottom-Gate Contact Activation

****

a) rca NO HF DIP (rca)

---

b) anneal 900C 15min (tubeB2 rec. 171 FF=15min)

****

20. Back-Gate Compensation Implant (splits)

****

stepper2 pattern CNC job WEI FILL align to back poly

---

b) coat (coater)

---

c) shoot (stepper2)

---

d) develop (developer)

---

e) counterdoping implants machine 6001 (Ion Implant Services)

---

f) piranah clean (pre-metal strip)

---

g) ash resist (asher)

****

21. Back-Gate Activation

****

a) rca NO HF DIP (rca)

---

b) anneal 600C (tubeB4 rec. 176 FF=1hr)

****

22. Deposit LTO Passivation

****

a) rca NO HF DIP (rca)

---

b) deposit 4k LTO (tubeA7 rec. 462 FF=40min)

****

23. Backside Wafer Strip

****

a) HMDS (HMDS)
---
b) coat (coater)
---
c) hardbake (developer program 82)
---
d) HF dip to remove 4kA LTO plus reox (oxide)
---
e) plasma etch (AME500 chamberB, BACKSIDE ETCH B)
---
f) BOE dip to make sure all clear (oxide)
---
g) ash resist (asher)

****
24. Contact Cuts
****
stepper2 pattern CC job WEI FILL align to poly
a) HMDS (HMDS)
---
b) coat (coater)
---
c) shoot (stepper2)
---
d) develop (developer)
---
e) oxide plasma etch (AME5000 chamberA, KEITH CD, modify etch step to 5sccm O2)
---
f) HF dip to open contact cuts (oxide)
---
g) ash resist (asher)

****
25. Metal Deposition
****
a) pre-metal clean (pre-metal)
---
b) Al-Si deposition (endura)

****
26. Metal Definition
****
stepper2 pattern CM job WEI FILL align to backgate
a) HMDS (HMDS)
---

b) coat (coater)

---

c) shoot (stepper2)

---

d) develop (developer)

---

e) hardbake (developer program 43 reprogram bake module 86 for 150C)

---

f) metal plasma etch (etcher-3 rec. 32)

---

g) dump rinse in panetch rinser to get rid of chlorine so resit will ash (panetch)

---

h) ash resist (asher)

---

i) dump rinse in panetch rinser (panetch)

---

j) metal sinter (tubeB8 rec. 710)

References

[1] B. Davari, H. Hovel, and G. G. Shahidi, "SOI Technology Outlook for Sub-0.25 μm CMOS, Challenges and Opportunities," *Proc.1993 IEEE Int. SOI Conf.*, pp. 4-5.

[2] J. Lasky, S. Stiffler, F. White, and J. Abernathy, "Silicon-On-Insulator (SOI) by Bonding and Etch-Back," *Int. Elec. Dev. Meeting (IEDM)*, pp. 684-687, 1985.

[3] K. Sakaguchi, et at., "ELTRAN by Splitting Porous Si Layers," *Proc. 9th ECS SOI Symp.*, vol. 99-3, p. 117, 1999.

[4] A. J. Auberton-Herve, "SOI: Materials to Systems," *Int. Elec. Dev. Meeting (IEDM)*, pp. 3-10, 1996.

[5] K. Izumi, M. Doken, and H. Ariyoshi, *Electronics Letters*, vol. 14, p. 593, 1978.

[6] J.-P. Colinge, *Silicon-on-Insulator Technology: Materials to VLSI.* Boston, MA: Kluwer Academic Publishers, 1991.

[7] J. Gautier and J. Sun, "On the Transient Operation of Partially-Depleted SOI NMOSFET's," *Elec. Dev. Letters*, vol. 16, no. 11, pp. 497-499,1995.

[8] A. G. Aipperspach, D. H. Allen, D. T Cox, N. V Phan, and S. N. Storino, "A 0.2-μm, 1.8 V, SOI, 550-MHz, 64-b PowerPC Microprocessor with Copper Interconnects," *IEEE Journal of Solid-State Circuits*, vol. 34, no. 11, pp. 1430-1435, 1999.

[9] G. G. Shahidi, et al., "Partially-Depleted SOI Technology for Digital Logic," *IEEE Int. Solid-State Circuits Conf.*, pp. 426-427, 1999.

[10] S. R. Wilson, T. Wetterroth, S. Hong, H. Shin, B. Y Hwang, M. Racanelli, J. Foerstner, M. Huang, and H. C. Shin, "Materials, Device and Gate Oxide Integrity Evaluation of SIMOX and Bonded SOI Wafers," *IEEE Int. SOI Conf.*, pp. 143-145, 1995.

[11] J. Tihanyi and H. Schlotterer, "Properties of ESFI MOS Transistors Due to the Floating Substrate and the Finite Volume," *IEEE Trans. on Electron Dev.*, vol. 22, p. 1017, 1975.

[12] C. Leroux, J. Gautier, A. J. Auberton-Herve, B. Giffard, M. Spalanzani, "Parasitic Transients Induced by Floating Substrate Effect and Bipolar Transistor on SOI Technologies," *Microelectronic Engineering 15*, pp. 199-202, 1991.

[13] S. S. Eaton and B. Lalevic, "Frequency Dependent Propagation Delay in Silicon-on-Sapphire Digital Integrated Circuits," *Solid-State Electronics*, vol. 21, pp. 1253-1257, 1978.

[14] S. S. Eaton and B. Lalevic, "The Effect of a Floating Substrate on the Operation of Silicon-

on-Sapphire Transistors," *IEEE Trans. on Electron Dev.*, vol. 25, no. 8, pp. 907-912, 1978.

[15] K. Kato and K. Taniguchi, "Numerical Analysis of Switching Characteristics in SOI MOSFET's," *IEEE Trans. on Electron Dev.*, vol. 33, no. 1, pp. 133-139, 1986.

[16] H. K. Lim and J. G. Fossum, "Transient Drain Current and Propagation Delay in SOI CMOS," *IEEE Trans. on Electron Dev.*, vol. 31, no. 9, pp. 1251-1258, 1984.

[17] D. Suh and J. G. Fossum, "Dynamic Floating-Body Instabilities in Partially-Depleted SOI CMOS Circuits," *Int. Electron Devices Meeting (IEDM) Tech. Dig.*, pp. 661-664, 1994.

[18] M. M. Pelella, J. G. Fossum, D. Suh, S. Krishnan, and K. Jenkins, "Low-Voltage Transient Bipolar Effect Induced by Dynamic Floating-Body Charging in PD/SOI MOSFET's," *Proc. 1995 IEEE Int. SOI Conf.*, pp. 8-9.

[19] A. Wei, "Measurement and Modeling of Transient Effects in Partially-Depleted SOI MOSFET's," S.M. Thesis, Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, June 1996.

[20] A. Wei, M. J. Sherony, and D. A. Antoniadis, "The Effect of Floating-Body Charge on SOI MOSFET Design," *IEEE Trans. on Electron Dev.*, vol. 45, no. 2, pp. 430-438, 1998.

[21] J.-P. Colinge, "Reduction of Kink Effect in Thin-Film SOI MOSFET's," *IEEE Electron Dev. Letters*, vol. 9, no. 2, pp. 97-99, 1988.

[22] H. Hazama, M. Yoshimi, M. Takahashi, S. Kambayashi, and H. Tango, "Suppression of Drain-Current Overshoot in SOI-MOSFET's Using Ultrathin SOI Substrate," *Electron. Lett.*, vol. 24, no. 20, p. 1266, 1988.

[23] H. L. Kim and J. G. Fossum, "A Charge-Based Large-Signal Model for Thin-Film SOI MOSFET's," *IEEE Trans. on Electron Dev.*, vol. 32, no. 2, pp. 446-457, 1985.

[24] L. Su, J. B. Jacobs, and D. A. Antoniadis, "Deep-Submicrometer Channel Design in Silicon-on-Insulator (SOI) MOSFET's," *IEEE Electron Dev. Letters*, vol. 15, no. 9, pp. 366-369, 1994.

[25] H. P. Wong, D. J. Frank, and P. M. Solomon, "Device Design Considerations for Double-Gate, Ground-Plane, and Single-Gated Ultra-Thin SOI MOSFET's at the 25 nm Channel Length Generation," *Int. Electron Devices Meeting (IEDM) Tech. Dig.*, pp. 407-410, 1998.

[26] M. Cao, T. Kamins, P. Vande Voorde, C. Diaz, and W. Greene, "0.18 μm Fully-Depleted Silicon-On-Insulator MOSFET's," *IEEE Electron Dev. Letters*, vol. 18, no. 6, pp. 251-253, 1997.

[27] I. Y. Yang, C. Vieri, A. Chandrakasan, and D. A. Antoniadis, "Back-Gated CMOS on SOIAS for Dynamic Threshold Voltage Control," *Int. Electron Devices Meeting (IEDM) Tech. Dig.*, pp. 877-880, 1995.

[28] *International Technology Roadmap for Semiconductors*, San Jose, CA: Semiconductor Industrial Association, 1999.

[29] Y. P. Tsividis, *Operation and Modeling of the MOS Transistor*. New York, NY: McGraw-Hill Publishing Company, 1987.

[30] R. F. Pierret, *Advanced Semiconductor Fundamentals*. Reading, MA: Addison-Wesley Publishing Company, 1987.

[31] J. A. del Alamo and R. M. Swanson, "Forward-Bias Tunneling: A Limitation to Bipolar Device Scaling," *IEEE Electron Dev. Letters*, vol. 7, no. 11, pp. 629-631, 1986.

[32] R. S. Muller and T. I. Kamins, *Device Electronics for Integrated Circuits*. New York, NY: John Wiley and Sons, 1977.

[33] G. A. M. Hurkx, D. B. M. Klaassen, and M. P. G. Knuvers, "A New Recombination Model for Device Simulation Including Tunneling," *IEEE Trans. on Electron Dev.*, vol. 39, no. 2, pp. 331-338, 1992.

[34] T. Ohno, M. Takahashi, A. Ohtaka, Y. Sakaibara, and T. Tsuchiya, "Suppression of the Parasitic Bipolar Effect in Ultra-Thin-FIlm NMOSFET's/SIMOX by Ar+ Ion Implantation into Source/Drain Regions," *Int. Electron Devices Meeting (IEDM) Tech. Dig.*, pp. 627-630, 1995.

[35] C. Hu, *Advanced MOS Device Physics: Hot Carrier Effects*, San Diego, CA: Academic Press, 1989.

[36] A. Wei, D. A. Antoniadis, and L. A. Bair, "Minimizing Floating-Body-Induced Threshold Voltage Variation in Partially-Depleted SOI CMOS," *IEEE Electron Dev. Letters*, vol. 17, no. 8, pp. 391-394, 1996.

[37] *MEDICI User's Manual*. Santa Clara, CA: Technology Modeling Associates, Inc., 1995.

[38] J. M. Rabaey, *Digital Integrated Circuits*. Upper Saddle River, NJ: Prentice Hall, 1996.

[39] A. G. Sabnis and J. T. Clemens, "Characterization of the Electron Mobility in the Inverted <100> Si Surface," *Int. Electron Devices Meeting (IEDM) Tech. Dig.*, pp. 18-21, 1979.

[40] Y. Taur and T. H. Ning, *Fundamentals of Modern VLSI Devices*, Cambridge, UK: Cambridge University Press, 1998.

[41] Z. Lee, M. B. McIlrath, and D. A. Antoniadis, "Two-Dimensional Doping Profile Characterization of MOSFET's by Inverse Modeling Using *I-V* Characteristics in the Subthreshold Region," *IEEE Trans. on Electron Dev.*, vol. 46, no. 8, pp. 1640-1649, 1999.

[42] M. Rodder, S. Aur, and I. C. Chen, "A Scaled 1.8 V, 0.18 μm Gate Length CMOS Technology: Device Design and Reliability Considerations," *Int. Electron Devices Meeting (IEDM) Tech. Dig.*, pp. 415-418, 1995.

[43] K. Mistry, T. Ghani, M. Armstrong, S. Tyagi, P. Pachan, S. Thompson, S. Yu, and M. Bohr, "Scalability Revisited: 100 nm PD-SOI Transistors and Implications for 50 nm Devices," *VLSI Symposium*, pp. 204-205, 2000.

[44] I. Y. Yang, "Study of Sub-0.5 μm SOI-with-Active-Substrate (SOIAS) Technology for Ultra-Low Power Applications," Ph.D. Thesis, Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, September 1996.

[45] S. M. Sze, *VLSI Technology*. New York, NY: McGraw-Hill Publishing Company, 1988.

[46] M. Martinez, "Chemical-Mechanical Polishing: Route to Global Planarization," *Solid State Tech.*, p. 26, May 1994.

[47] S. Bengtsson, "Semiconductor Wafer Bonding: A Review of Interfacial Properties and Applications," *Journal of Elec. Materials*, vol. 21, no. 8, pp. 841-862, 1992.

[48] *Model 6EC Laboratory Planarizer User's Manual*, San Luis Obispo, CA: Strasbaugh, 1996.

[49] T. H. Smith, "Device Independent Process Control of Dielectric Chemical Mechanical Polishing," Ph.D. Thesis, Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, September 1999.

[50] M. A. Schmidt, "Silicon Wafer Bonding for Micromechanical Devices," *Solid-State Sensor and Actuator Workshop*, pp. 127-131, 1994.

[51] W. Kern, "The Evolution of Silicon Wafer Cleaning Technology," *Semiconductor Cleaning Technology*, p. 3, 1989.

[52] Andrew Loomis, Lincoln Laboratories, personal correspondence.

[53] H. Seidel, L. Csepregi, A. Heuberger, and H. Baumgaertel, "Anisotropic Etching of Crystalline Silicon in Alkaline Solutions," *Journal of the Electrochemical Society*, vol. 137, no. 11, pp. 3612-3632, 1990.

[54] A. Merlos, M. Acero, M. H. Bao, J. Bausells, and J. Esteve, "TMAH/IPA Anisotropic Etching Characteristics," *Sensors and Actuators*, pp. 737-743, 1993.

[55] G. Brown and A. Revesz, "Defect and Bulk Electrical Conduction in SIMOX Buried Oxide," *IEEE Int. SOI Conf.*, pp. 32-33, 1993.

[56] J. H. Lee, G. Taraschi, A. Wei, T. A. Langdo, E. A. Fitzgerald, and D. A. Antoniadis, "Super Self-Aligned Double-Gate (SSDG) MOSFET's Using Oxidation Rate Difference and Selective Epitaxy," *Int. Electron Devices Meeting (IEDM) Tech. Dig.*, pp. 71-74, 1999.

[57] H. S. P. Wong, K. K. Chan, and Y. Taur, "Self-Aligned (Top and Bottom) Double-Gate MOSFET with a 25 nm Thick Silicon Channel," *Int. Electron Devices Meeting (IEDM) Tech. Dig.*, pp. 427-430, 1997.

[58] X. Huang, W. C. Lee, C. Kuo, D. Hisamoto, L. Chang, J. Kedzierski, E. Anderson, H. Takeuchi, Y. K. Choi, K. Asano, V. Subramanian, T. J. King, J. Bokor, and C. Hu, "Sub 50-nm FinFET: PMOS," *Int. Electron Devices Meeting (IEDM) Tech. Dig.*, pp. 67-70, 1999.

[59] J. M. Hergenrother, et al., "The Vertical Replacement-Gate (VRG) MOSFET: A 50-nm Vertical MOSFET with Lithography-Independent Gate Length," *Int. Electron Devices Meeting (IEDM) Tech. Dig.*, pp. 75-78, 1999.

[60] A. Ritenour, "Fabrication Technology for Double-Gate Field Effect Transistors," S. M. Thesis, Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, May 1999.