2010

High field induced stress suppression of GIDL effects in TFTs

Andrew McCabe

HIGH FIELD INDUCED STRESS SUPPRESSION OF GIDL EFFECTS IN TFTS

By

Andrew M. McCabe

A Thesis Submitted

In Partial Fulfillment

of the Requirements for the Degree of

Master of Science

in

Microelectronic Engineering

Approved by:

Prof: _____

Dr. Karl D. Hirschman (Thesis Advisor)

Prof: _____

Dr. Sean L. Rommel (Committee Member)

Prof: _____

Dr. Santosh K. Kurinec (Committee Member)

Prof: _____

Dr. Robert E. Pearson (Program Director)

Prof: _____

Dr. Sohail A. Dianat (Department Head)

DEPARTMENT OF ELECTRICAL AND MICROELECTRONIC ENGINEERING
KATE GLEASON COLLEGE OF ENGINEERING
ROCHESTER INSTITUTE OF TECHNOLOGY
ROCHESTER, NEW YORK
JULY 2010

HIGH FIELD INDUCED STRESS SUPPRESSION OF GIDL EFFECTS IN TFTS

By

Andrew M. McCabe

_____	_____
Andrew M. McCabe	July 29, 2010

ABSTRACT

Gate-Induced drain leakage (GIDL) is an unwanted short-channel effect that occurs at higher drain biases in an overdriven off state of a transistor. The GIDL is the result of a deep depletion region that forms in the drain at high drain-to-gate biases. The depletion region causes significant band bending which in-turn allows conductive band-to-band tunneling creating excess current. In a PFET, electrons tunnel from the drain to the body while holes tunnel into the drain.

By utilizing the effects of high energy, or "hot", electrons, the GIDL current in an accumulation-mode PFET can be suppressed. This suppression is thought to be due to local creation of interface charge at the gate-oxide/silicon interface located close to the drain end of the transistor. This charge in-turn creates a mirror charge in the silicon, which acts like a pseudo-asymmetrical lightly-doped drain structure. Up until now, this effect has only been demonstrated on the first order. The goal of this study is to further investigate the effects of high-field stress on the suppression of GIDL in accumulation-mode PFETs. An overview of background information, simulations, fabrication, electrical characterization, and physical characterization are presented in this study.

TABLE OF CONTENTS

LIST OF FIGURES

ACKNOWLEDGEMENTS

I would like to thank everyone who has helped me complete this project. I would like to thank my advisor, Dr. Karl Hirschman without his support I would not have been able to complete this work. I would also like to thank my committee members Dr. Sean Rommel and Dr. Santosh Kurinec for guidance in completing this project. I would also like to thank the members of Team Eagle: Ryan Rettmann, Chris Shea, and Patricia Meller for their help with both process and interpretation of the data. I would like to thank Dr. Surenda Gupta of the mechanical engineering department for instruction in using the Park Systems XE150 AFM along teaching me the theory behind AFM. I gratefully acknowledge the support provided by the National Science Foundation for funding the acquisition of the PSIA XE-150 Atomic Force Microscope under the DLR grant # EEC-530575. I would also like to extend special thanks to Silvaco for the academic donation of Silvaco/Simucad TCAD/EDA tools that were used in this study. I would like to thank my project sponsors Dr. Carlo Kosik-Williams and Dr. Greg J. Couilliard from Corning, Inc. and Dr. Timothy Treadwell from Carestream, Inc.; without their support I would not have been able to get this project off of the ground. I would like to thank Krystina Sylvester for her moral support during my graduate studies. Last but not least, I would like to thank my parents Michael and Marcianna McCabe for their support of all of my endeavors throughout my college career.

CHAPTER 1 - INTRODUCTION

1.1 Introduction and Motivation

As of recently the use of thin-film transistors (TFTs) have proliferated both the integrated circuit (SOI technology) and display industries. Although TFTs offer several advantages over conventional bulk metal-oxide-semiconductor field-effect transistors (MOSFETs), TFTs still suffer from problems that are also observed in comparable bulk devices; one of which is gate-induced drain leakage (GIDL). This study focuses on gate-induced drain leakage (GIDL), which occurs due to band-to-band tunneling at the drain end of the transistor. GIDL can cause an increase in the off-state current of a transistor by multiple orders of magnitude, depending on bias conditions. With TFTs being used in many mobile (low-power) applications the current draw of a transistor in the off-state becomes very important. A high GIDL current can increase power consumption and reduce battery life.

By utilizing the effects of high-field stress on accumulation-mode p-channel TFTs, the GIDL current has been reduced by up to two orders of magnitude. This reduction of GIDL is proposed to be localized charge centers within 200 nm of the drain end of the transistor. An example current-voltage (I-V) characteristic of a device pre-stress and post-stress is shown in Fig. 1. The transistor tested was fabricated on a SmartCut^{TM} SOI substrate using a self-aligned gate and a thermally grown gate oxide. The applied stress condition was V_G = 15 V and V_D = -6 V, with a resulting gate-drain electric field of 4.2 MV/cm.

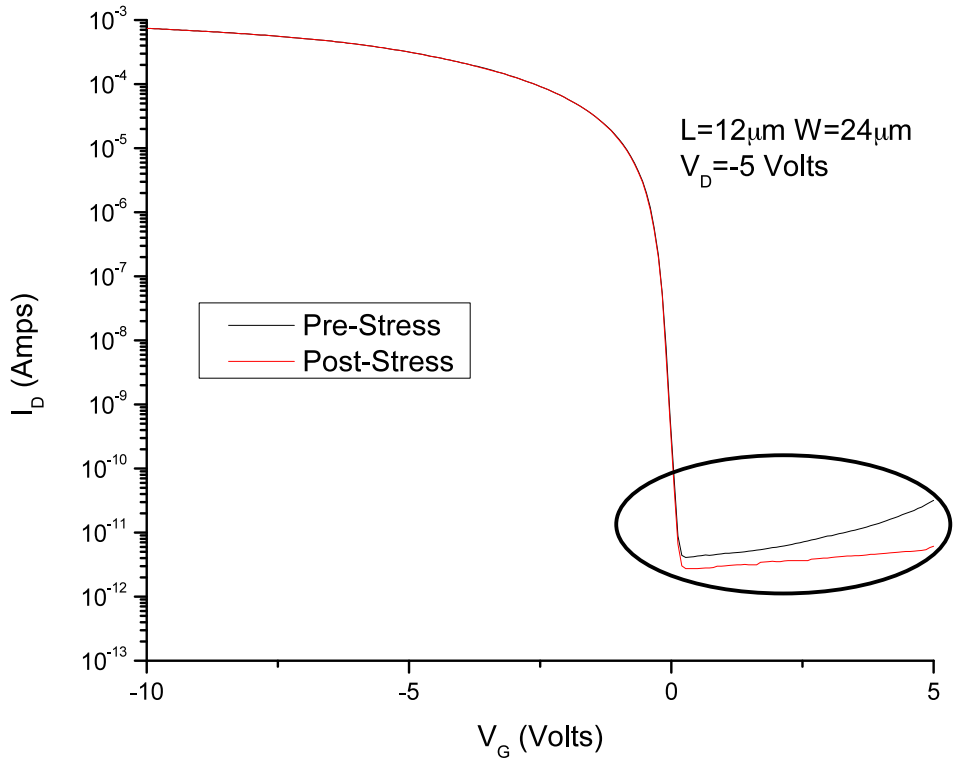

Fig. 1. Example accumulation-mode p-type TFT I_D-V_G characteristic pre-stress and post-stress showing reduction in GIDL current.

1.2 Corporate Partnerships

Since 2005 RIT has had a supported research partnership with Corning Inc. to help to develop their silicon on glass (SiOG) substrate technology. In the past year, RIT has also developed a partnership with Carestream Health Inc. to help develop amorphous silicon based x-ray detectors on Corning's SiOG substrate for biomedical applications. The properties of the SiOG substrate do not allow for conventional CMOS processing as seen in the traditional high performance CMOS IC industry. A developed low-temperature CMOS TFT process results in both enhancement-mode NMOS and PMOS devices in only four mask levels while keeping the maximum processing temperature less than or

equal to 600ºC, which satisfies the temperature constraints of the EAGLE$^{2000®}$ display glass [1].

1.3 Work Covered by this Study

The primary focus of this work was to investigate the effects of high-field stress, specifically between the gate and drain, on the off-state characteristics of an accumulation-mode PFET TFT through fabrication, testing and simulation. A background discussion on GIDL is provided, followed by simulation results using Silvaco Atlas® [2] which demonstrates the physics behind the GIDL suppression mechanism. The details of the low-temperature TFT fabrication process are presented along with results from fabricated devices tested both pre-stress and post-stress under various stress-bias treatment combinations. Electrostatic Force Microscopy (EFM) has been explored as a potential technique to directly measure charge levels that account for the GIDL suppression, with results supporting a consistent model of GIDL suppression and device operation.

CHAPTER 2 – GATE-INDUCED DRAIN LEAKAGE

2.1 Gate-Induced Drain Leakage

The GIDL observed in bulk silicon MOSFETs is caused by band-to-band tunneling in the gate-to-drain overlap region [3-5]. The band-to-band tunneling can be attributed to the fact that a deep depletion region forms where the gate overlaps the drain with increasing drain bias while the gate is at 0 V bias or in an over-driven off-state. In a NFET, electrons pass into the drain and holes (minority carriers) travel from the drain to the channel through carrier tunneling, thus creating an excess current. The effect of GIDL on the I_D-V_G characteristics of a NFET is shown in Fig. 2.

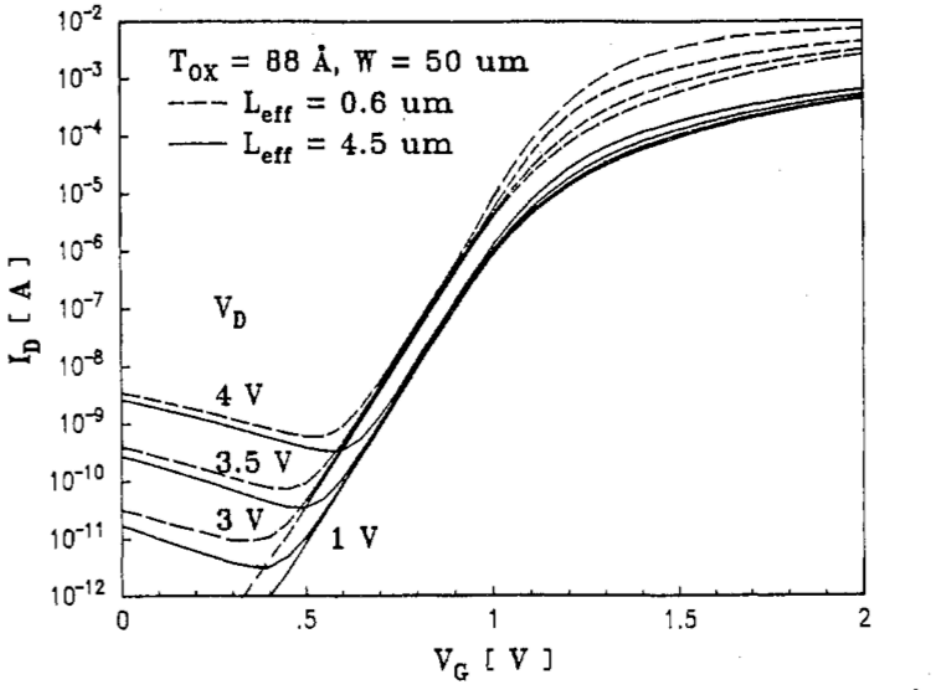

Fig. 2. Subthreshold characteristics of two NFETs showing GIDL current at various drain voltages. As the drain voltage increases so does the current in the off-state [4].

The model that is most commonly used for GIDL is the model that was proposed by Chen *et al.* in 1987; which is referred to as the constant band-bending model. The constant band-bending model is assumes a constant level of band-bending required for band-to-band tunneling (GIDL) to occur. The graphical representation of GIDL and the band bending at the drain edge of the channel is shown in Fig. 3.

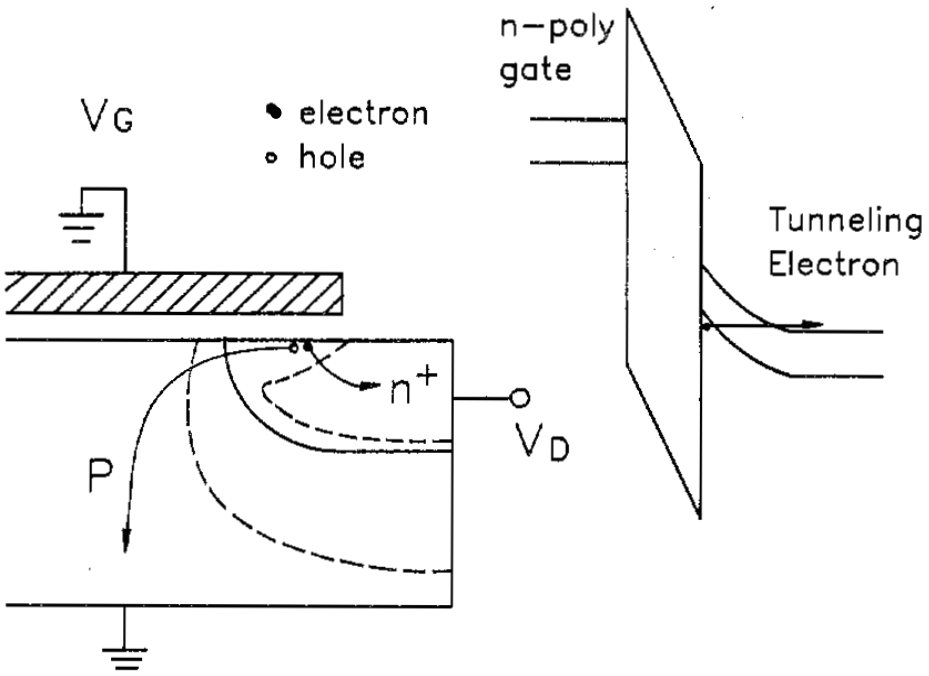

Fig. 3. GIDL in a gated diode configuration showing deep-depletion region and hole and electron band-to-band tunneling. Energy band diagram showing tunneling election from valence to conduction band resulting in excess drain current [4].

The band-to-band tunneling current will be the greatest in the region where the electric field is highest. For tunneling to occur, the band bending must exceed the bandgap of the material which creates available states for the tunneling process. In this model, the band-bending is assumed to be a constant value of 1.2 eV which is the minimum value for band-to-band tunneling to occur in silicon [4]. The simplified equation for determining

the maximum surface electric field in the gate drain overlap region, derived by Chen *et al.*, is shown in (1).

$$E_s \approx \frac{V_{dg} - 1.2}{3T_{ox}} \tag{1}$$

In (1) E_s is the electric field at the surface, V_{dg} is the drain voltage with respect to the gate, T_{ox} is the oxide thickness, 1.2 *eV* is the minimum value of band-bending for band-to-band tunneling to occur, and 3 is the ratio of the permittivity of silicon (Si) to the permittivity of the silicon-dioxide (SiO_2) gate insulator. From the value of E_s determined in (1) the GIDL current value can be calculated by applying basic band-to-band tunneling theory as shown in (2) [3, 4].

$$I_D = AE_s exp(-B/E_s) \tag{2}$$

In (2) A is a constant that can be determined experimentally while the equation for B is shown in (3). For the units to be correct in (2), A must have the units of cm/ohm.

$$B = \frac{\pi m^{*1/2} E_g^{3/2}}{2\sqrt{2} q\hbar} \tag{3}$$

The assumption of $m^* = 0.2m_o$ yields a value of $B = 21.3$ MV/cm [4].

2.2 GIDL in SOI Based Devices

When examining GIDL that is apparent in SOI devices, the conventional theory of GIDL in bulk devices must be modified to accommodate the structural difference of a TFT. A TFT is electrically isolated from the substrate while a bulk device is connected to the substrate. In a TFT the GIDL current has no substrate contact for the excess carriers to dissipate, which causes the body potential to rise to the point at which the body-source junction can become forward biased. This initiates electron injection from

the n^+ source into the body, which travels to the drain resulting in excess GIDL current. This amplification is a demonstration of parasitic bipolar junction transistor (PBT) operation [6], and is analogous to the on-state "kink effect" observed on transistors built on isolated substrates [7]. A visual representation of GIDL in a N-channel TFT is shown in Fig. 4.

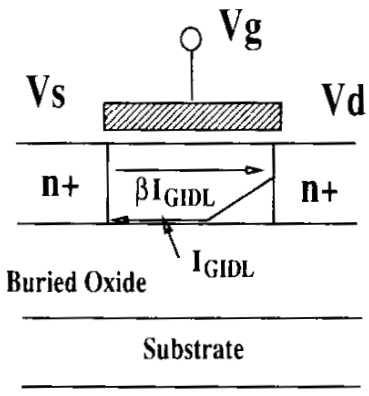

Fig. 4. Schematic of N-channel TFT showing the PBT GIDL effect [6].

The gain (β) of the PBT is a function of the channel length of the transistor. For long channel devices $\beta = 0$ but as the channel length decreases the β of the PBT increases [6]. An example of the short channel GIDL behavior compared with the long channel GIDL behavior is shown in Fig. 5. The plot shows a drain sweep while the gate is being held at a constant voltage. The drain sweep shows both the onset and magnitude of GIDL. As L decreases, the magnitude of GIDL increases by the gain of the PBT ($\beta+1$).

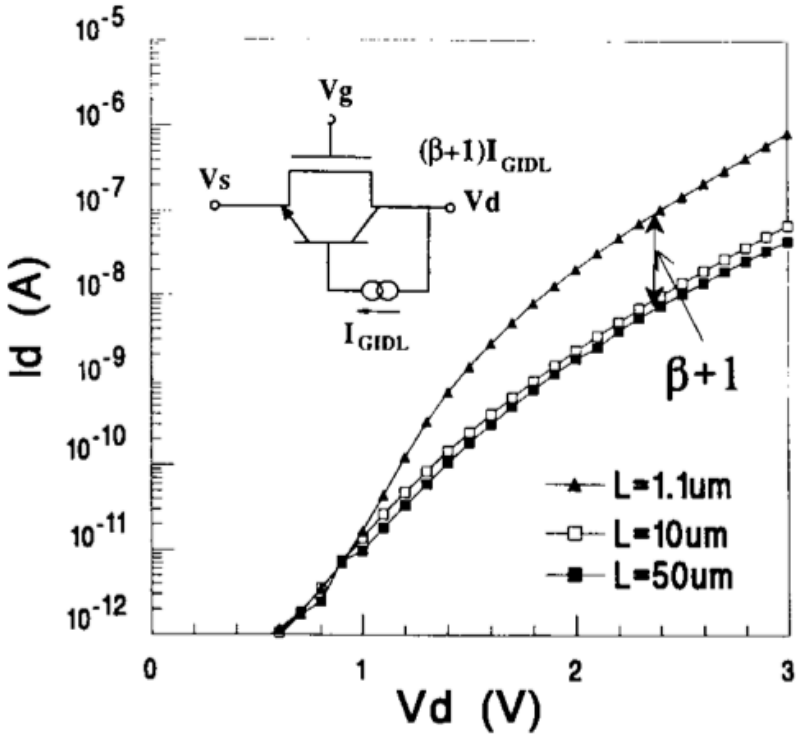

Fig. 5. Drain sweep comparison of SOI TFT short-channel and long-channel GIDL behavior that shows the effects of a parasitic bipolar transistor.

2.3 Conventional Methodology for Reducing GIDL

Since GIDL is due to band-to-band tunneling of carriers at the drain edge, a reduction in band-bending will result in a reduction or full suppression of GIDL. Lightly-doped drain (LDD) regions can be used to modify the field distribution at the drain end of the device [8, 9]. A modern LDD structure consists of a n^- region next to the drain of an NFET or a p^- region next to the drain of a PFET. LDDs can be implemented in both bulk and SOI technologies to reduce the severity of band-bending at the drain edge. The reason for using a small LDD region instead of having the entire drain lightly-doped is because of series resistance and transistor drive current considerations. There are a few ways to implement an LDD each with its own benefits and drawbacks. The gate can either fully

overlap or partially overlap the LDD region [5, 8, 9], with both strategies shown in Fig. 6. Theoretically, the LDD with oxide spacer will have the latest onset of GIDL with respect to drain voltage, but since the gate has less control over the LDD region the on-state current will be less because of the series resistance of the LDD region. The polysilicon spacer LDD process [8] should have more GIDL current because of the fact that the gate can still influence the surface potential at the LDD-drain interface but will also have a higher drive current because of the extra gate control.

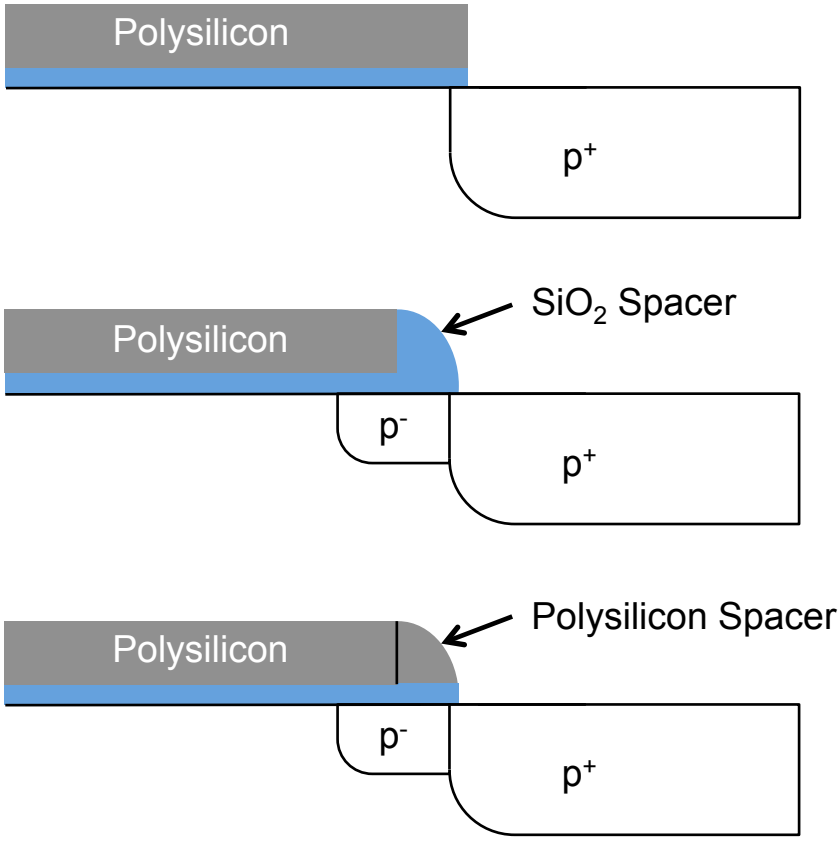

Fig. 6. From top to bottom: No LDD, fully overlapped LDD (with polysilicon spacer), and partially overlapped LDD oxide spacer. Adapted from [9].

There are two main process factors that influence the characteristics of a LDD. The dose of the lightly-doped region governs the series resistance of the device and the amount of band-bending at both the LDD to channel edge and LDD to drain edge. The polysilicon

re-oxidation thickness governs the field the drain edge will see and then in-turn controls the amount of band bending and band-to-band tunneling [5]. Kurimoto *et al.* shows that there are in fact two regions in the LDD for GIDL current generation. Fig. 7 shows a cross section of the LDD MOSFET with the two current generation regions. Region 1 corresponds to the region at the corner of the LDD/channel edge, whereas region II corresponds to the gate/drain overlap region. Where the GIDL is concentrated is dependant on the dose of the LDD. If LDD dose is low the GIDL will be more concentrated at the gate edge because the electric field will be more concentrated there as the channel causes some depletion of the LDD.

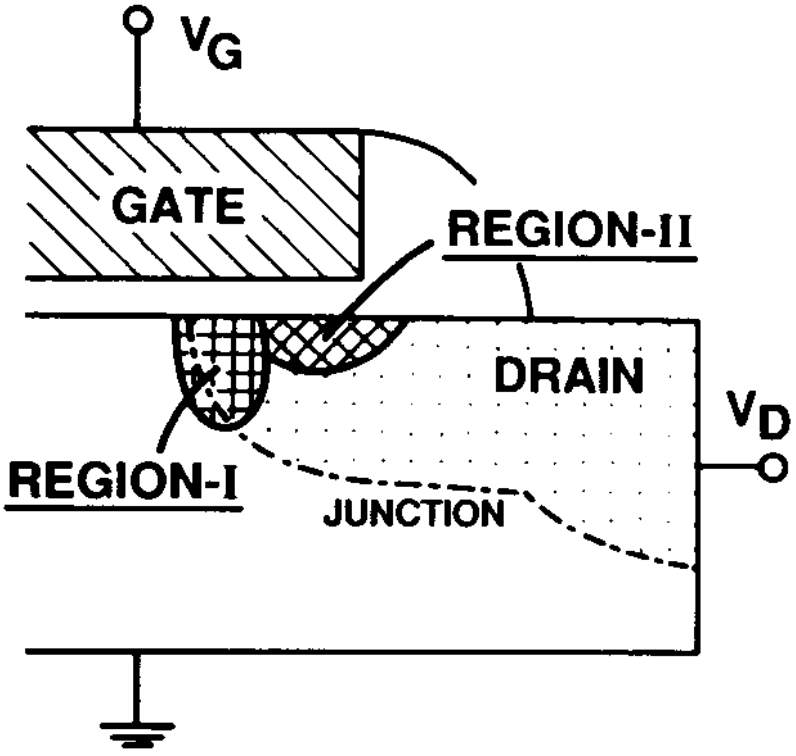

Fig. 7. Cross section of LDD MOSFET with two GIDL current generation regions shown [5].

Another way to suppress GIDL on SOI devices is reverse mode stress, where a stress voltage is applied to the source and the drain is grounded, with a gate voltage slightly above threshold. Reverse mode stress in an SOI TFT can induce interface traps at the source end of the channel due to high energy electrons, which decreases the gain of the PBT thus decreasing GIDL current [10]. Fig. 8 shows a representation of the SOI TFT after reverse mode stress where the traps that are created are denoted by the '*x*'s at the interfaces.

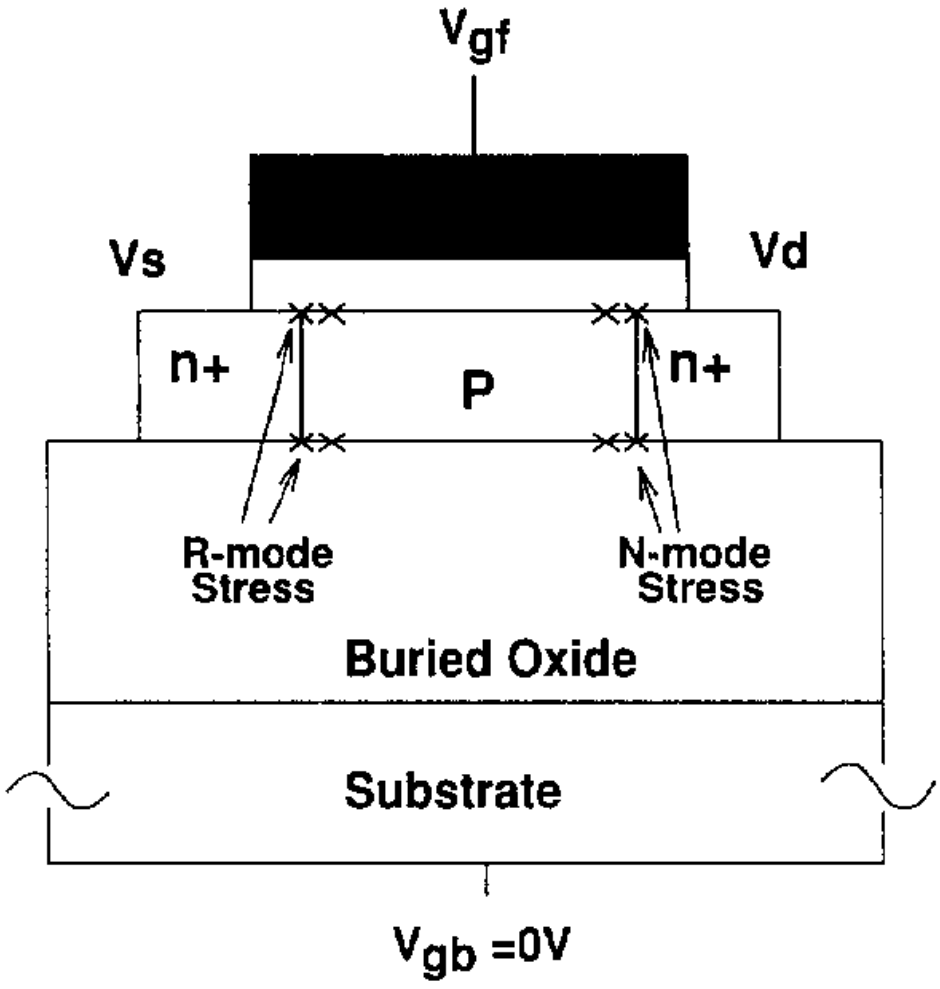

Fig. 8. SOI TFT showing interface traps created after either normal mode (N-mode) or reverse mode (R-mode) hot carrier stressing [10].

2.5 Extraction of Effective Charge in Oxide after High Field Stress

It is possible to extract the effective amount of charge per unit area in the oxide after a given stress condition [11]. The charge extraction technique follows traditional GIDL theory presented in [4] and (2) in section 2.1. To extract the charge amount a change in surface field due to the amount of charge must be defined. The change in surface field due to high field stress is

$$\Delta E_s = \int_0^{t_{ox}} \frac{t_{ox} - x}{\epsilon_{Si} t_{ox}} Q(x) dx \equiv \frac{Q_{eff}}{\epsilon_{Si}} \tag{4}$$

\where t_{ox} is the oxide thickness, *Q(x)* is the volumetric charge density in the oxide, and ε_{Si} is the dielectric constant of silicon [11]. The change in surface field is due solely to the amount of charge in the entire oxide so that is why the integral goes throughout the entire thickness. The location of '0' is defined at the SiO_2/Si interface. Now knowing the entire surface field (E_s+ΔE_s) the theory provided in (2) can be used to solve for the GIDL current. The solution for the GIDL current is shown in (5).

$$I_d(post - stress) = A(E_s + \Delta E_s) exp\left(-\frac{B}{E_s + \Delta E_s}\right) \tag{5}$$

By assuming that ΔE_s is small compared to E_s and using a Taylor's Expansion (5) can be rearranged as shown in (6) [11].

$$\begin{aligned} I_d(post - stress) &= AE_s(1 + \Delta E_s/E_s) exp(-B/E_s) exp\left(\frac{B}{E_s^2}\Delta E_s\right) \\ &\approx I_d(pre - stress) exp\left(\frac{BQ_{eff}}{\epsilon_{Si} E_s^2}\right) \end{aligned} \tag{6}$$

By being able to relate the post-stress current to both the effective charge in the oxide and pre-stress current the effective oxide charge can be extracted by a simple rearrangement

of (6). The resulting equation for effective charge in the oxide in terms of pre-stress GIDL, post-stress GIDL, and the field between the gate and drain (buried in E_s) is shown in (7) [11].

$$Q_{eff} = \frac{\epsilon_{Si} E_s^2}{B} ln \left(\frac{I_d(post - stress)}{I_d(pre - stress)} \right) \tag{7}$$

2.6 High Field Stress Induced Effects on GIDL

Nearly all work published on GIDL has focused on NFET devices, with investigations on the application of electric field stress showing changes in GIDL behavior due to hot (energetic) electrons [10]. While this investigation focuses on the accumulation-mode PFET the proposed mechanism of GIDL suppression is quite similar, also involving hot electrons. Note that complimentary applications that would require hot-hole carriers to serve in a similar role may not be possible due to lower amount of kinetic energy gained under the same applied stress conditions [12].

While high-field stress on electronic devices is usually associated with the creation of defects, in certain situations the resulting changes in I-V behavior can be most interesting and even have potential benefits. The goal of this thesis is to understand the effects of high field stress on GIDL suppression observed on accumulation-mode PFETs, which has been accomplished through simulation, fabrication and device testing. The next chapter describes electrical simulation of GIDL behavior, providing insight on the mechanism of GIDL suppression related to changes in the physical device structure.

CHAPTER 3 – THE ACCUMULATION-MODE PFET

3.1 PMOS TFT Operation

TFTs fabricated on SiOG and have been made in p-type boron doped films. Given the temperature limitations of the SiOG substrate, conventional isolation techniques cannot be used to create inversion-mode CMOS TFTs. An inversion-mode NFET and accumulation-mode PFET are utilized to realize CMOS without doping compensation. The operation of the accumulation-mode PFET is quite different than that of an inversion mode PFET. The threshold voltage of the device is at the onset of accumulation where the surface potential is equal to the flat-band voltage. On a P-type substrate a negative surface potential will attract (accumulate) hole carriers to the surface. While most of the current conduction occurs at the surface of the device, some of the current also travels through the body since this device is without metallurgical junctions. Under a low gate bias (subthreshold) the body of the transistor is partially depleted meaning that body current can still flow. When the bias gets more positive (less negative), the body of the transistor becomes fully depleted effectively shutting the transistor off. The different modes of operation of the accumulation mode PFET are shown in Fig. 9.

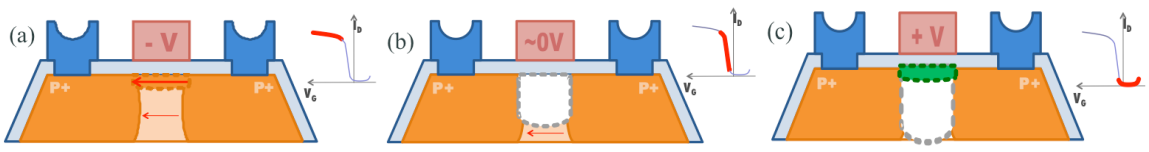

Fig. 9. (a) Accumulation mode PFET above threshold showing both current in the accumulation layer as well as current in the body. (b) Accumulation mode PFET in subthreshold regime showing body conduction. (c) Accumulation mode PFET in the OFF-state showing inversion layer at the surface of the device [13].

While TFTs can have a body contact, they normally consist of only three terminals instead of four. The doping in the initial silicon layer used for this study is relatively low (~10^{15} cm^{-3}) so that the device body will be fully depleted (emptied of majority carriers) and free of the floating body effects found in partially depleted TFTs with no body contact [14]. A graphical comparison between a fully-depleted and a partially-depleted TFT is shown in Fig. 10.

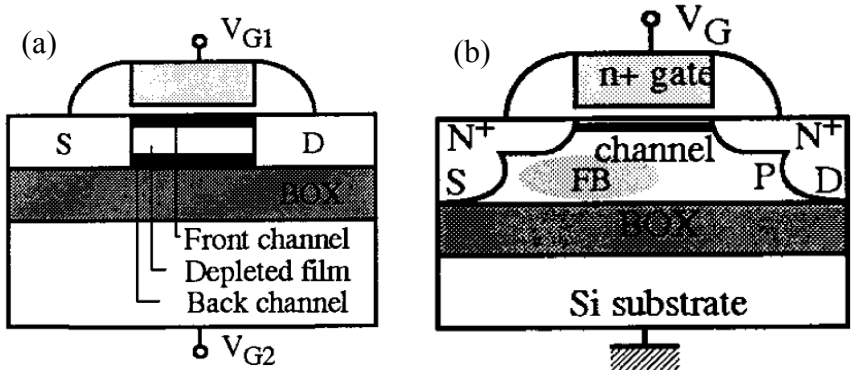

Fig. 10. (a) Cross-section of fully-depleted SOI TFT showing both top and back gate. (b) Cross-section of partially depleted SOI TFT showing the floating body [14].

3.2 Low-Temperature CMOS TFT Process

Devices in this study are fabricated using a developed low temperature CMOS TFT process [1], omitting steps that are not required for the PFET. The process keeps thermal processes to 600ºC or below; thermal constraints set forth by the glass substrates create many process integration challenges. At 600 ºC it is difficult to incorporate the dopants into substitutional sites in the lattice. To achieve a high concentration of electrically active boron, lattice amorphization and regrowth through solid phase epitaxy (SPE) was used. Following amorphization, at about 500 ºC to 600 ºC the silicon regrows from a seed crystal resulting in incorporation of the dopant into substitutional lattice sites

(activation). A co-implant of fluorine and boron was used to achieve amorphization in the source/drain regions. Results have shown that in these implanted regions the crystal regrows from the top down, resulting in activation levels of greater than 90% of the total amount of dopant in the silicon [15].

A traditional thermally grown gate oxide cannot be used on SiOG substrates since thermal oxidation requires ~ 900 ºC to occur. The developed process used a low-pressure chemical vapor deposition (LPCVD) low-temperature oxide (LTO), deposited at 425 ºC in a vacuum tube furnace. The target thickness for the LTO gate oxide was 50 nm, with a charge level of less than 10^{11} cm^{-2}. Since SOI wafers were also used in this experiment, both a thermally grown gate oxide in a dry O_2 ambient and the deposited LTO were explored.

Since the capability of in-situ doped polysilicon at temperatures below 600 ºC was not available, a metal gate was used instead. Molybdenum was selected because it has a work function of 4.53 eV and because it is a refractory metal with a high melting point. A molybdenum gate also allows for a self-aligned process similar to what is seen in the high performance IC industry today. Fluorine and boron co-implants were done to form the source/drain regions, followed by the 600 ºC SPE process. LTO was deposited as an inter-level dielectric, and contact openings were patterned and etched. Aluminum was deposited and patterned for the electrodes and interconnects, with a final anneal (sinter) at 425 ºC in forming gas (5% H_2 in N_2).

3.3 Modeling the Accumulation-Mode PFET

Modeling of the device was done in Silvaco Data Systems' Atlas package [2]. Atlas is a package that allows for modeling of a variety of semiconductor devices. Previous work has been done to implement various mobility and band-bending models to match the electrical characteristics of devices that have been fabricated at RIT [16]. The model for the accumulation-mode PFET incorporates both concentration and field dependent numerical mobility models and band-to-band tunneling (GIDL). Unfortunately Atlas does not have the model presented in Chapter 2 for GIDL but the model it provides is sufficient. The model used is based on a similar exponential relationship with three different fitting parameters. The equation for the carrier generation due to band-to-band tunneling is shown in (8) where *BB.A*, *BB.GAMMA*, and *BB.B* are user-definable tunneling parameters and *E* is the local electric field [2]. The parameters used for this study were a *BB.A* of 2 x 10^{15}, a *BB.GAMMA* of 2, and a *BB.B* of 8x10^{6} V/cm.

$$G_{BBT} = BB.A(E^{BB.GAMMA})exp\left(-\frac{BB.B}{E}\right) \tag{8}$$

Both of these models have been iteratively calibrated for the devices fabricated at RIT [16]. The models also incorporate a 1x10^{11} cm^{-2} charge density across the gate-oxide/silicon interface to match both capacitance-voltage (C-V) and current-voltage (I-V) data historically measured on the fabricated devices. The as modeled accumulation-mode PFET subthreshold characteristic is shown in Fig. 11.

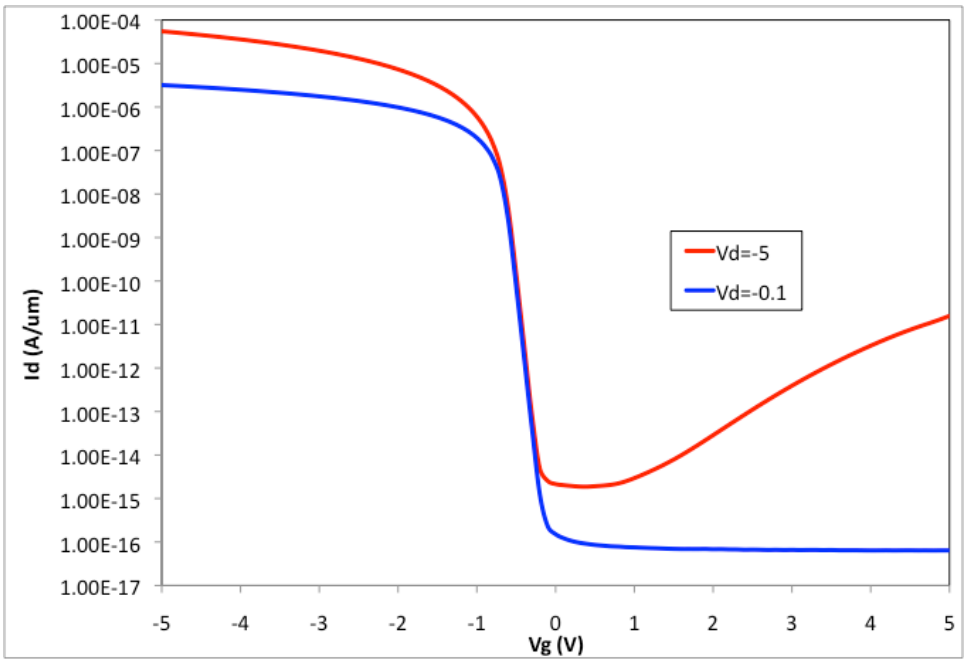

Fig. 11. Modeled subthreshold characteristics of an accumulation mode PFET with L=10 µm using Atlas with an implementation of band-to-band tunneling model for realization of GIDL.

3.4 Modeling of the High Field Stress

High field stress on the accumulation-mode PFET was simulated using the device structure shown in Fig. 12, with appropriate physical models. In this study the hot-carrier model was used, which is based on the lucky electron model [17] for hot-carrier injection into a gate oxide. Once invoked, all the models that have been specified are continuously solved over the structure mesh. Overall the mesh has relatively coarse spacing except near the drain region of the transistor where the bias is highest.

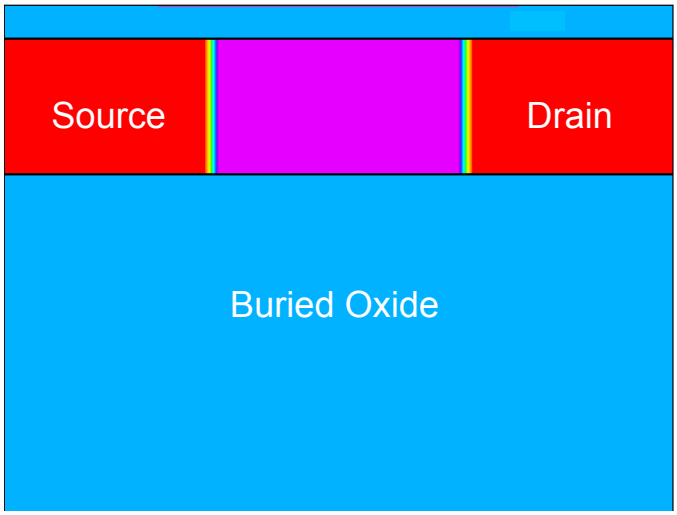

Fig. 12. Atlas cross-section of the simulated accumulation-mode PFET. Source/drain regions were specified with doping concentrations of 10^{19} cm^{-3} with a Gaussian lateral distribution of 0.05 µm. Ideal electrodes are specified and have no physical dimension.

Since Atlas cannot count the amount of charge injected over time into the gate oxide, modeling of the stress condition is only to give an idea of where the charge is actually being injected into the oxide. Once the location is known, the amount of charge in that region of the oxide can be varied to examine its influence on the transistor characteristics. The results from modeling a high-stress condition on a 10 µm channel length accumulation-mode PFET are shown in Fig. 13.

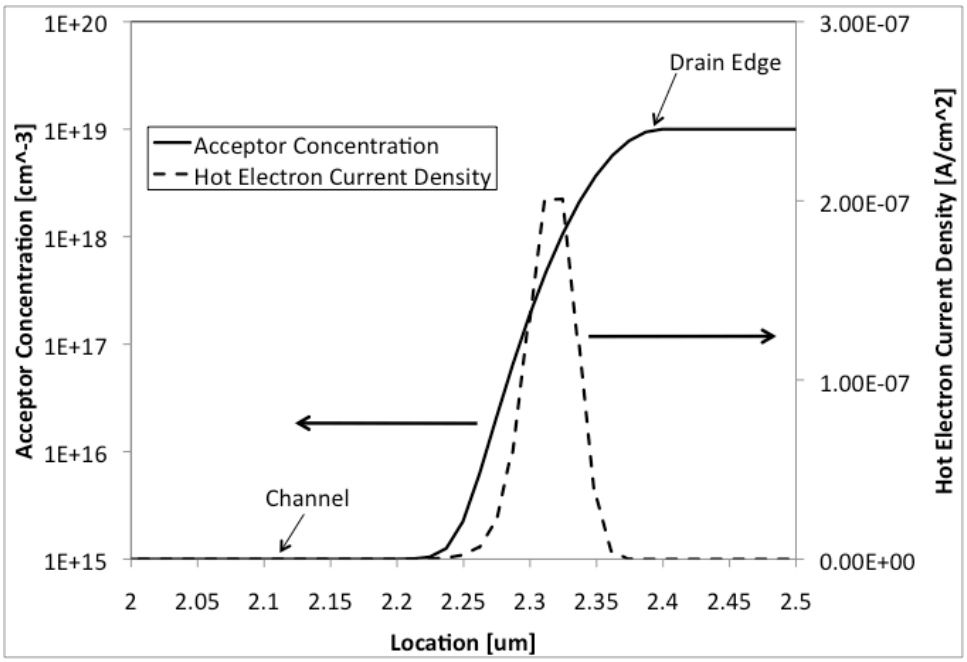

Fig. 13. Cutline of Atlas modeled high-field stress close to the Si/SiO_2 interface, showing the location of the hot-electron current with respect to the acceptor concentration. The location of the hot electron current is within ~200 nm of the drain edge.

Once the hot electron current region was established, interface charge was placed around this location of high probability, which was confined close to the drain edge. The interface charge density was varied to explore the effects of an equivalent oxide charge on the performance of the transistor. The excess interface charge was specified in a box from the drain edge that extends 200 nm into the body. The influence of the distance of this extension on the simulation of GIDL was found to be dependent on parameters in the band-to-band tunneling model, and the lateral characteristic of the doping profile. The lateral characteristic of the doping profile determines the lateral distribution of dopants beyond the specified drain end; a lateral characteristic of zero denotes a perfectly abrupt junction. The saturation-mode (V_{DS}=-5 V) subthreshold characteristics of the various charge levels are shown in Fig. 14.

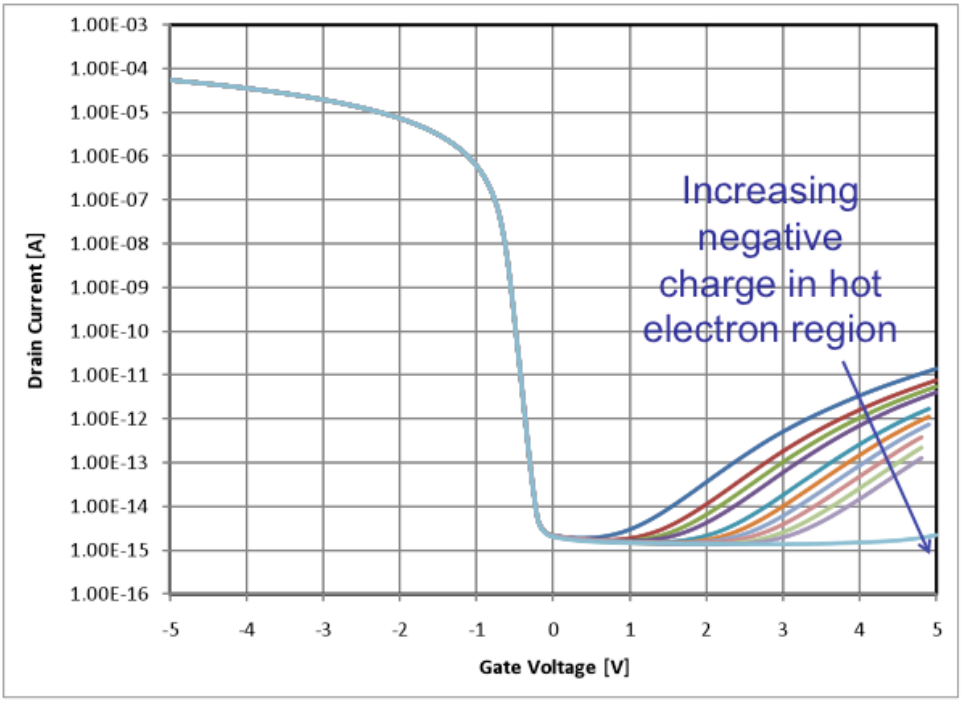

Fig. 14. Saturation-mode (V_{DS}=-5 V) subthreshold characteristics of various levels of charge starting with the virgin device then increasing charge level in increments of 10^{10} cm^{-2} up to 10^{11} cm^{-2}, with only the last curve showing 10^{12} cm^{-2} of charge. The last curve demonstrates suppression of GIDL at 5 V on the gate.

At a charge level of approximately 10^{12} cm^{-2} the subthreshold characteristics show almost no GIDL at 5 V on the gate. Although there has been charge introduced at the drain-edge of the transistor there is no visible decrease in drive current as with LDDs. There is also no shift in the extrapolated threshold voltage of the transistor. Since the accumulation-mode PFET is turned off by inverting the channel (electrons), an examination of the amount of inversion charge can provide insight into how the interface charge influences the transistor. The amount of inversion charge (electron concentration) at the silicon surface and at the drain-end of the channel with V_{GS} = 5 V and V_{DS} = -5 V is shown in Fig. 15.

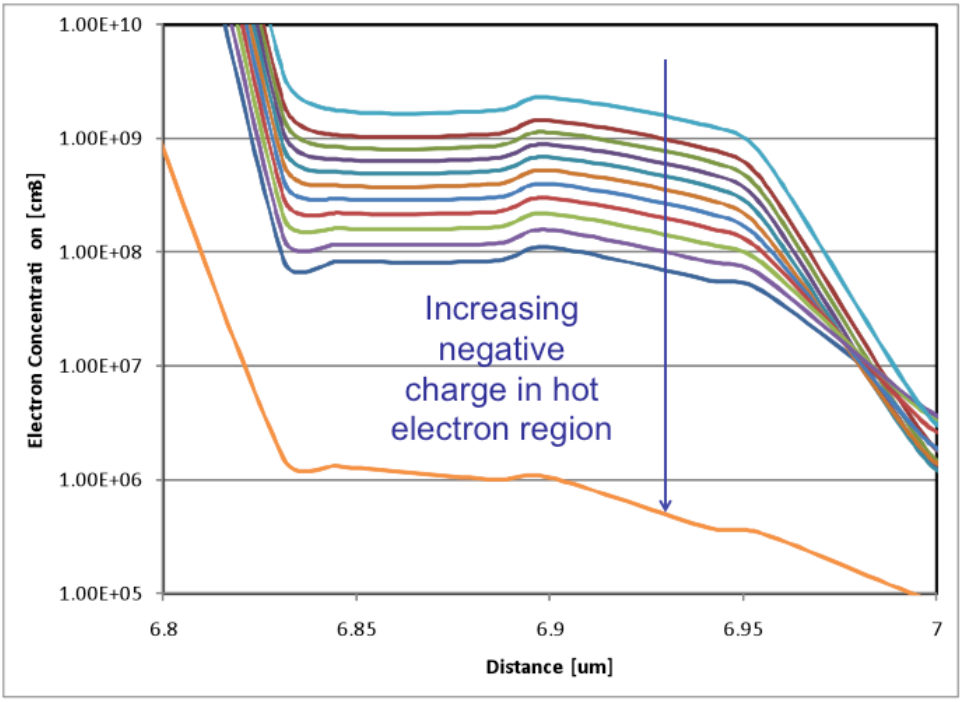

Fig. 15. Electron concentration at the drain end of the channel (V_G=5 V and V_D=-5 V) showing that with an increasing level of negative interface charge in the hot-electron region there are less electrons in the corresponding inversion layer. Interface charge level increases by 10^{10} cm^{-2} with every data series. The series with the lowest concentration has an interface charge level of 10^{12} cm^{-2}. The virgin device is the topmost curve.

With increasing negative charge at the gate-oxide/silicon interface the amount of inversion charge in the silicon decreases, which results in a decrease in the electric field gradient across the drain into the channel and less band-to-band tunneling (GIDL). The energy band diagrams between the drain and channel of the transistor comparing three different charge levels are shown in Fig. 16.

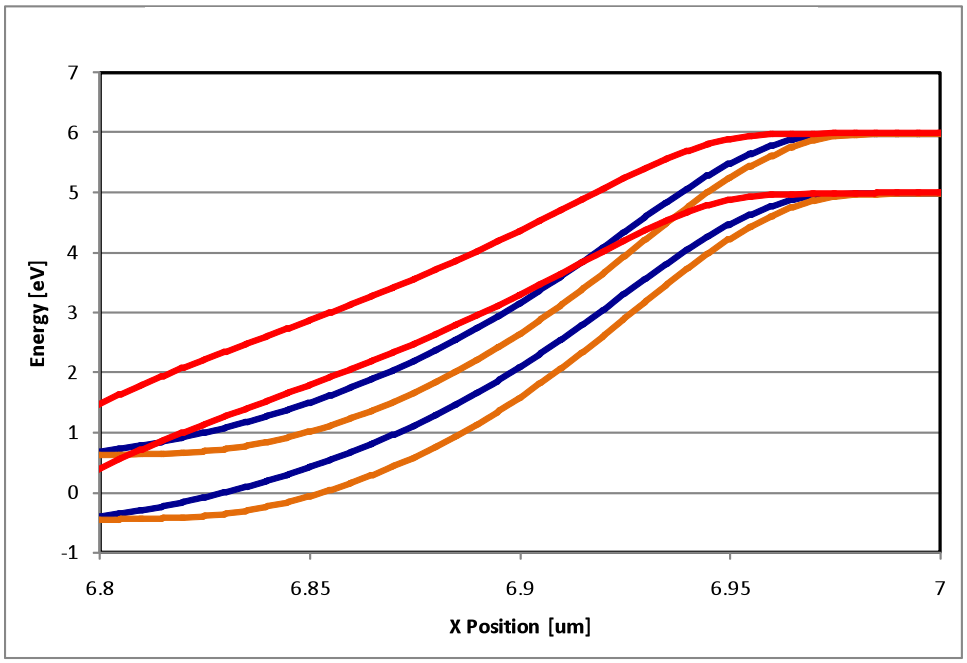

Fig. 16. Energy band diagrams showing band bending at the drain end of the transistor with V_D=-5 V and V_G=5 V. The red curve represents a charge level of -2x10^{12} cm^{-2} while the blue curve represents a charge level of +1x10^{11} cm^{-2} and the orange curve represent a charge level of +1x10^{12} cm^{-2}. Note that positive charge increases the band bending and decreases the energy barrier width, thus promoting GIDL.

The amount of band bending at the drain edge of the transistor decreases with increasing amount of negative charge in the hot-electron region of the gate-oxide. The maximum electric field can be found by taking the maximum slope of the bands. Once the maximum electric field is found, the tunneling probability can be calculated using (9) where m* is the electron effective mass, E_g is the energy gap and E_{max} is the maximum electric field in the device [7].

$$T_t \approx exp\left(-\frac{4\sqrt{2m^*}E_g^{3/2}}{3q\hbar E_{max}}\right) \tag{9}$$

The results from using (14) with various interface charge levels in the gate-oxide are shown in Fig. 17. The tunneling probability was normalized to the probability at which there is no charge in the hot electron region of the oxide.

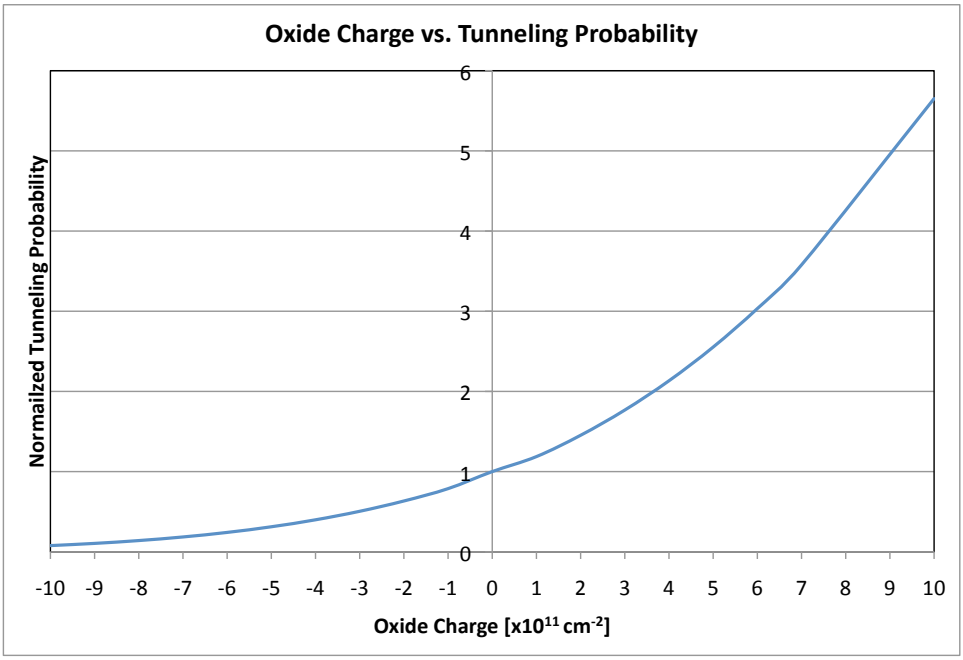

Fig. 17. Normalized tunneling probability vs. amount of charge in the hot electron region of the oxide. The reference is the tunneling probability with zero added charge.

As the amount of negative interface charge increases the tunneling probability approaches zero (GIDL suppression), while if the charge becomes less negative, or more positive, the tunneling probability increases rapidly.

CHAPTER 4 – ELECTRICAL CHARACTERIZATION

4.1 PFET Fabrication and Operation

Accumulation-mode PFETs were fabricated on both SOI and SiOG substrates, using the procedure outlined in chapter 3. As specified, the SOI devices had a thermally grown gate oxide, whereas the SiOG devices had a gate oxide deposited using LPCVD. Example characteristics in the linear and saturation operating regions shown on a linear scale from one of the SOI devices is shown in Fig. 18.

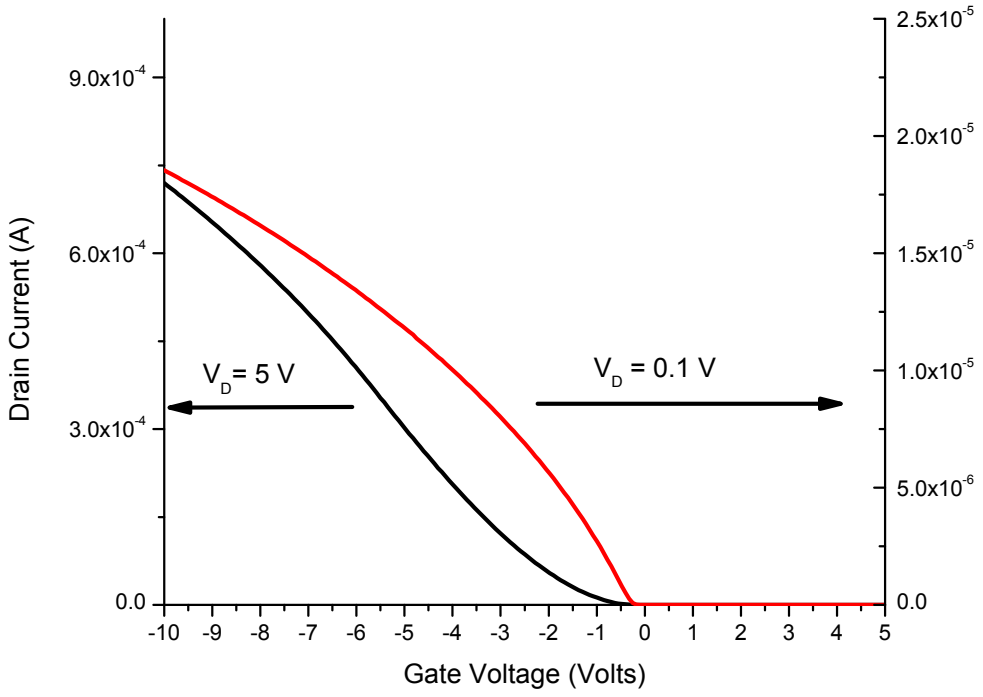

Fig. 18. Example linear ($|V_D|$ = 0.1 V) and saturation ($|V_D|$ = 5 V) mode characteristics of fabricated accumulation-mode PFETs on SOI.

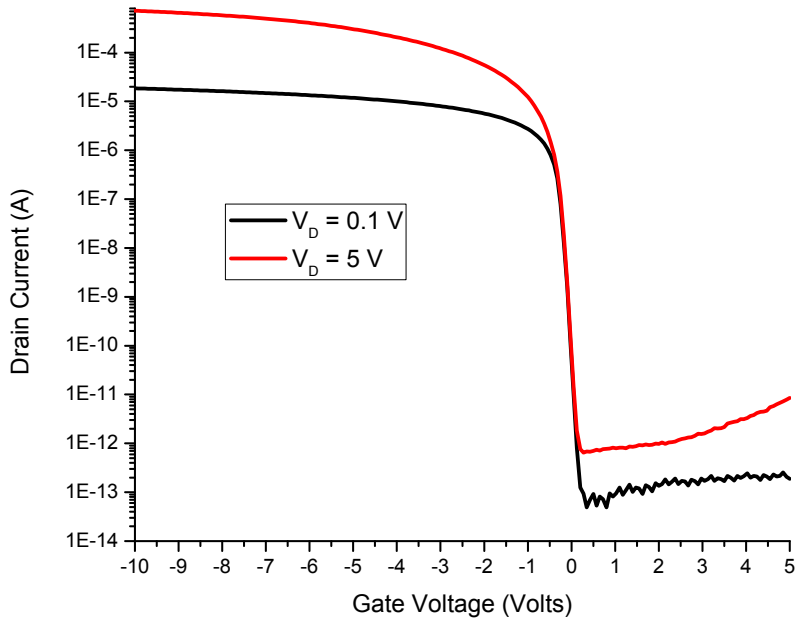

Fig. 19. Example linear ($|V_D|$ = 0.1 V) and saturation ($|V_D|$ = 5 V) mode characteristics of fabricated AMPFETs on SOI plotted on a log scale.

As expected, the process yielded enhancement-mode transistors with a threshold voltage magnitude of ~0.3 V; close to the theoretical value for this transistor design. For examination of the off-state leakage current, Fig. 19 shows the linear and saturation characteristics on a log(I_D) scale. The saturation mode sweep shows a higher leakage current in the off-state which is expected due to the higher electric field at the drain end of the transistor which gives rise to GIDL.

4.2 Demonstration of Induced GIDL Suppression

Representative pre-stress and post-stress characteristics from both SOI and SiOG devices are shown in Fig. 20. The stress voltages for the initial characterization were +15 V on the gate and -15 V on the drain. Both the SOI and SiOG wafers show similar

characteristics, including the bias stress-induced GIDL suppression phenomenon. The degraded subthreshold swing and on/off state performance of the SiOG device relative to that of the SOI device is explained elsewhere by Manley *et al.* [18].

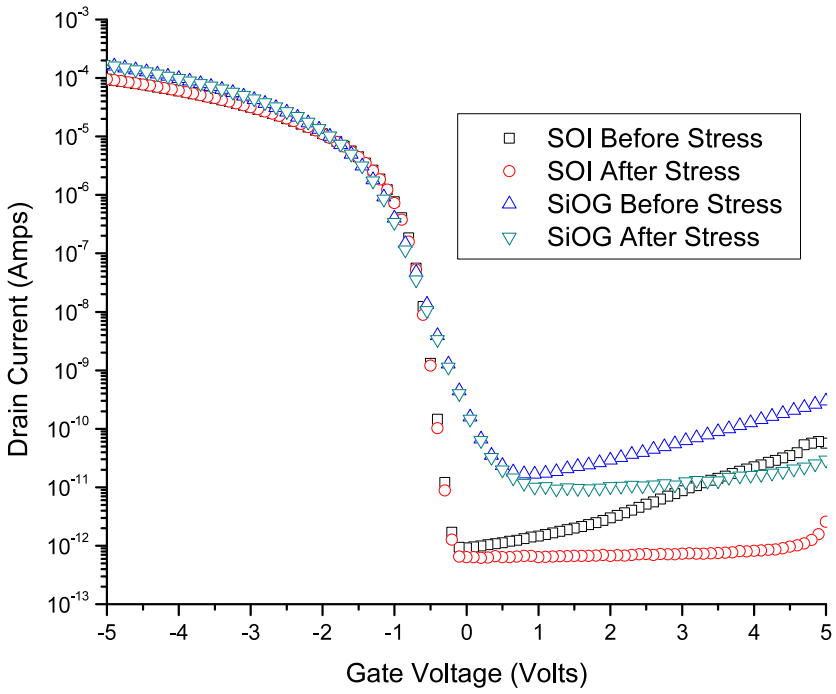

Fig. 20. Representative pre-stress and post-stress transfer characteristics from AMPFETs fabricated on SOI (thermal gate oxide) and SiOG (LPCVD gate oxide) substrates. W/L=24 µm/12 µm and V_D=-5 V. The mechanism of GIDL suppression does not appear to be dependent on the substrate material.

4.4 Stress-Bias Experiment

To determine the effects of the stress voltage on both the gate and the drain, a 5x5 full factorial experiment was performed on SOI substrates using a thermally grown gate oxide. This experiment involved all combinations of the following voltages applied to the gate (positive) and drain (negative) electrodes: 3, 6, 9, 12, and 15 V. Each treatment

combination was examined using two different techniques. A conventional gate-voltage sweep with a constant drain voltage (I_D-V_G) was performed to examine the off-state behavior of the transistor both pre-stress and post-stress. A drain-voltage sweep with a constant gate voltage was performed to examine change in the onset and magnitude of GIDL with each stress condition. Fig. 21 shows example measurements taken pre-stress and post-stress on SOI devices.

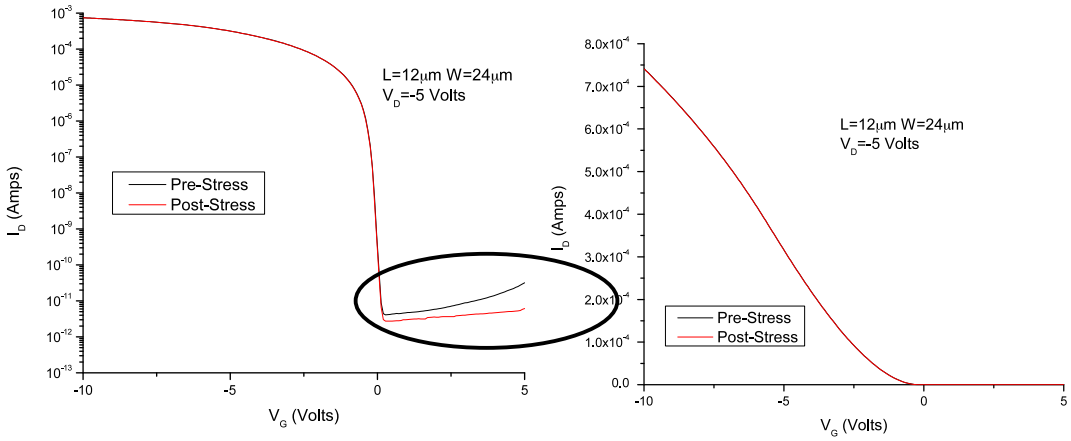

Fig. 21. Pre-stress and post-stress overlay of I_D-V_G Characteristics of W/L=24 μm/12 μm transistor. This device was stressed at V_G=6 V and V_D=-12 V for 30 sec.

The plots in Fig. 21 show typical behavior, both before and after an applied stress condition. While there is a leakage current reduction (log scale), there is no reduction in the magnitude of the on-state current (linear scale). It should be noted that an applied stress field that is too high results in a threshold voltage shift. Fig. 22 shows the drain-voltage sweep of the device associated with the results in Fig. 21, with the gate in an over-driven off-state. The drain sweep is a common method used to characterize the onset and magnitude of GIDL current [3]. The drain sweep shows that once the gate-to-drain field is sufficient, band-to-band tunneling (GIDL) current becomes significant.

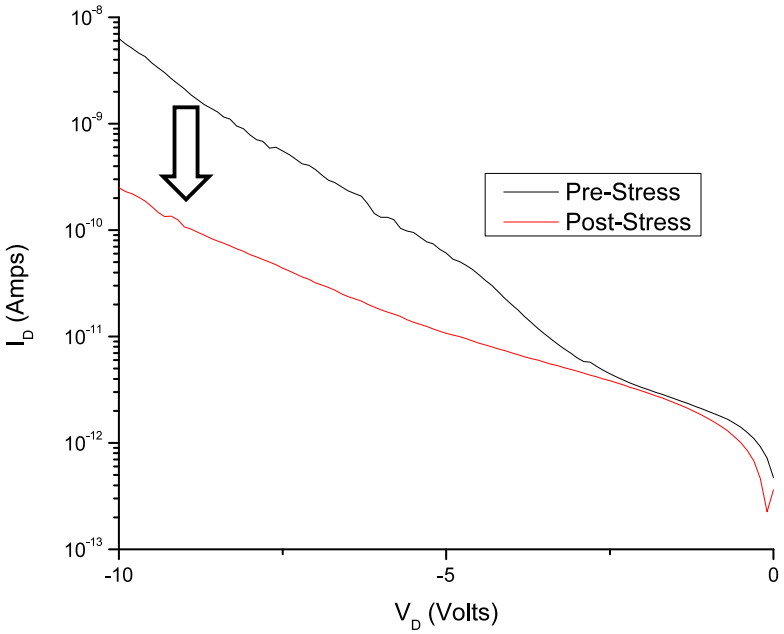

Fig. 22. Drain sweep of W/L=24 µm/12 µm transistor both pre and post-stress. This device was stressed at V_G=6 V and V_D=-12 V for 30 sec. V_G=5 V to show GIDL behavior.

The drain sweep measurement is a good method to look at the onset of GIDL and how the current behaves in higher bias conditions. Ratios of the post-stress to pre-stress current at V_D = -10 V and V_G = 5 V were calculated to compare the different treatment combinations. By calculating ratios at a higher bias there is less noise from the measurement setup itself. Fig. 23 shows a contour plot of the post-stress to pre-stress ratio against both the applied gate and drain stress voltages. All stress treatments were for a time of 30 sec.

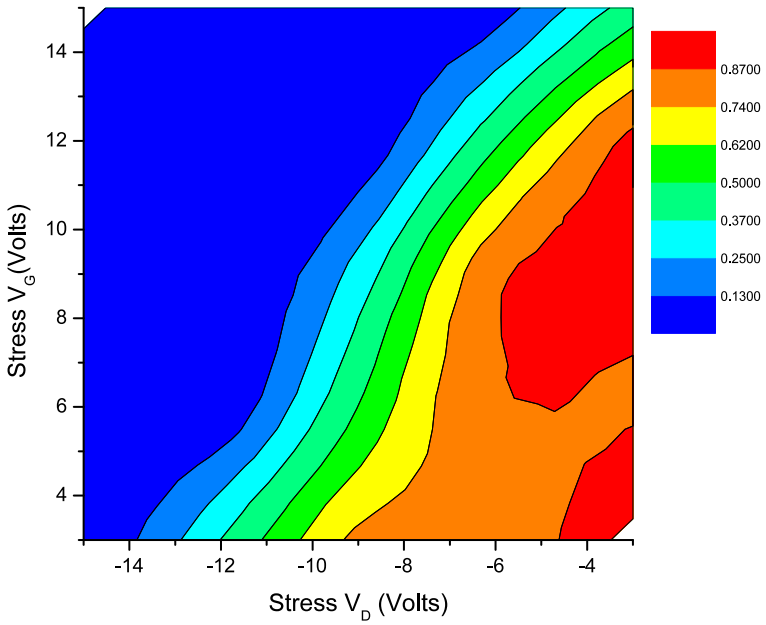

Fig. 23. Contour plot of I_D(post-stress)/I_D(pre-stress) measured at V_D=-10 V and V_G=5 V.

The contour plot shows that once a certain gate-to-drain field is reached, the GIDL suppression reaches a maximum (dark blue on the chart). Representation of the suppression vs. total gate-to-drain stress field is shown in Fig. 24.

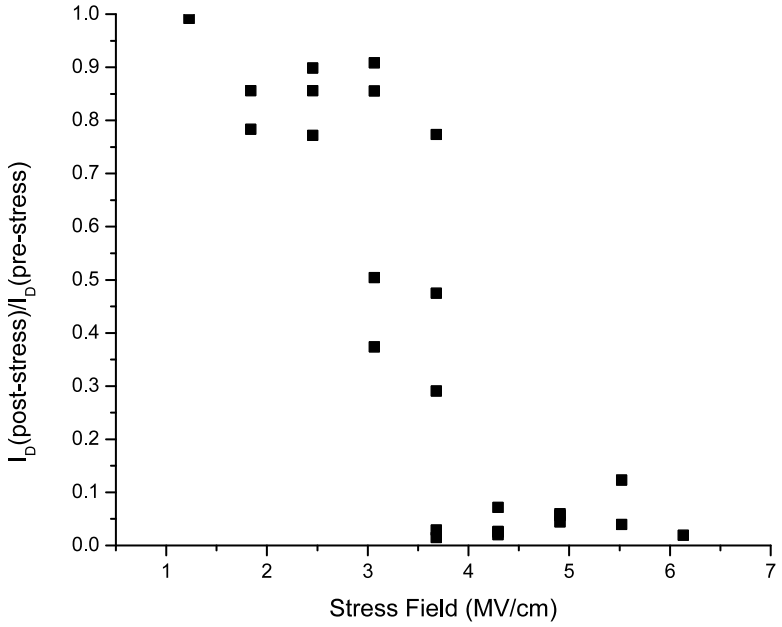

Fig. 24. Relationship between I_D(post-stress)/I_D(pre-stress) and stress field across the gate-oxide.

The post-stress current was measured to be up to two orders of magnitude lower than the pre-stress current depending on the stress condition. While several treatment combinations resulted in an applied field of 3.7 MV/cm, it was the high V_{DS} / low V_{GS} combination that demonstrated the greatest suppression of leakage current. Once the field surpasses a stress field of approximately 4 MV/cm the transistor shows a considerable decrease in leakage current across all combinations.

Using a method presented in [11], the effective charge injected into the gate-oxide can be calculated. The derivation models a contribution to the surface voltage at the drain end due to an effective oxide charge. The results from this calculation are shown in Fig. 25. The injected charge density is calculated from the amount of GIDL suppression

to be approximately $1x10^{12}$ cm^{-2} for applied stress conditions that exceed 4 MV/cm. This value is consistent with simulations using the Atlas TCAD package (see Fig. 14).

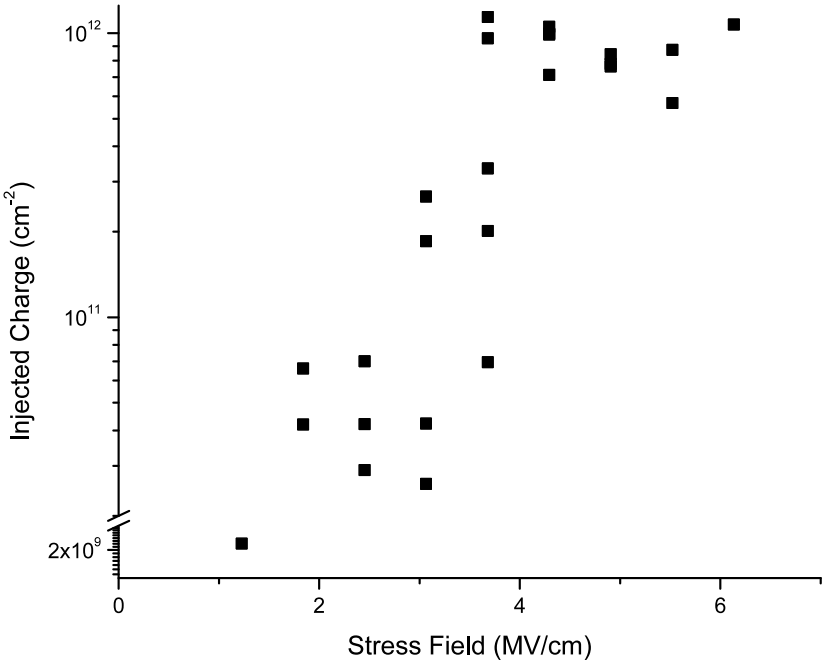

Fig. 25. Calculated injected charge density vs. stress field. Note the discontinuity in the y-axis.

4.4 Stress-Time Experiment

Up to this point a stress time of 30 sec was used to stress all of the devices. Fig. 26 shows the results of an experiment where stress time was varied to explore the relationship between stress time and GIDL suppression. For this experiment a 4.2 MV/cm stress condition was chosen. The reason a stress just beyond the minimum field that gives the maximum leakage suppression was chosen is to show if the device is not

stressed for a long enough time, the maximum suppression will be a result. In the figure as the stress time increases from 1 to 30 sec the ratio of post-stress to pre-stress current decreases showing that overall stress time does play a factor.

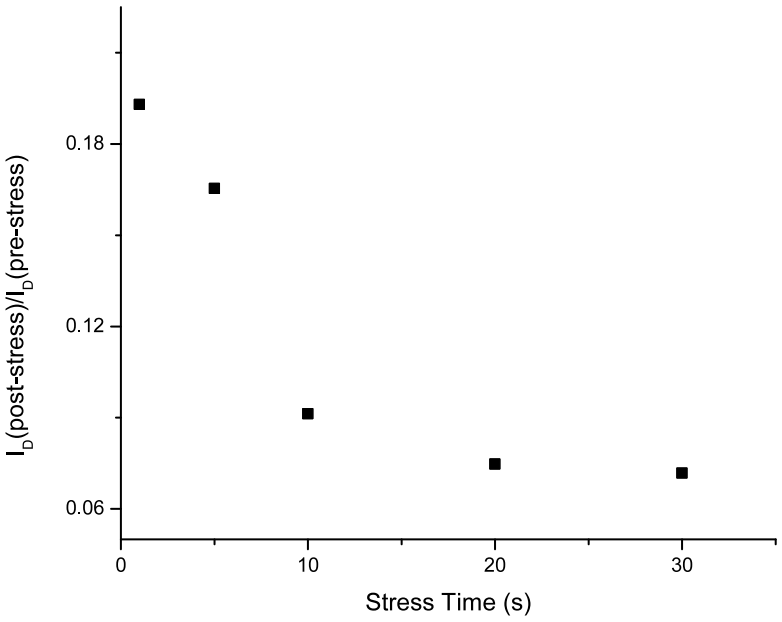

Fig. 26. Plot showing the relationship between GIDL suppression and time under stress. The applied stress voltages were V_G=15 V and V_D=-6 V.

During the stressing of the device a gate current of 1-10 nA was observed. If the calculations are carried through to calculate an state density over an area of 200 nm x 24 μm the amount of charge that flowed into the gate was anywhere between $3.9x10^{18}$ cm^{-2} and $3.9x10^{19}$ cm^{-2}. The large amount of charge flowing into the gate is indicative of the creation of interface states near that drain edge rather than a fraction of the gate current getting caught in the oxide.

4.5 Stress Stability

To examine the stability of the GIDL suppression five devices were stressed and repeatedly measured over a long time. The treatment combinations that were chosen were that showed the maximum suppression for each gate voltage with the minimum overall field between the gate and drain. These applied stress treatment combinations are as follows: V_G = 15 V V_D = -6 V, V_G = 12 V V_D = -9 V, V_G = 9 V V_D = -12 V, V_G = 6 V V_D = -12 V, and V_G = 3 V V_D = -15 V. The results are shown in Fig. 27. The discrepancy early on in the testing is due to changes in the test setup that influenced the overall noise floor.

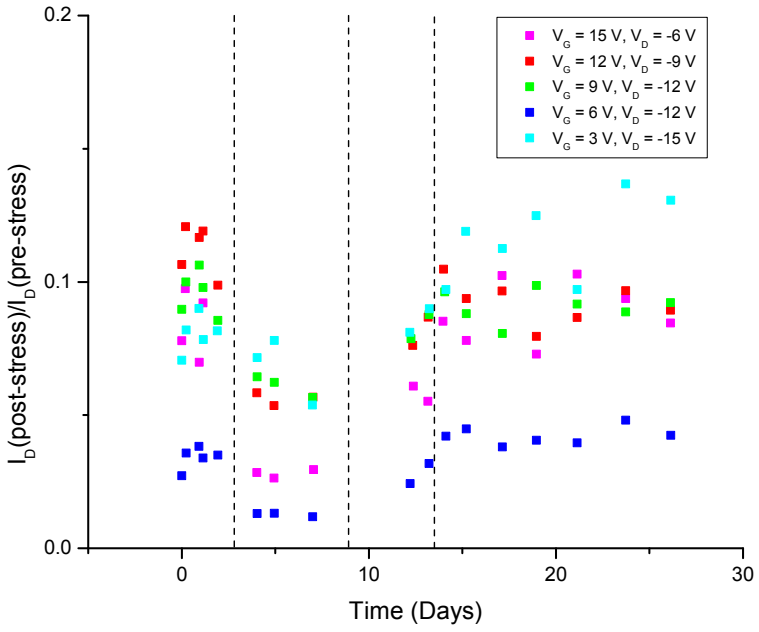

Fig. 27. I_D(post-stress)/I_D(pre-stress) measurements versus time after initial stress for various stress conditions. The ratio of post-stress to pre-stress is calculated at V_G = 4.8 V and V_D = -5 V. Dashed lines represent changes in the test setup that influence the overall noise floor of the measurement.

The results show that even several weeks after the initial stress all of the stressed transistors are still showing GIDL suppression. The stability of the suppression is proposed to be due to the dynamic filling of the interface traps during the measurement that are generated at the time of stress, which makes it appear that the GIDL suppression is permanent. This theory is supported by measurement data that shows a sharp decreasing slope of the drain current at the start of the measurement. An example of this data is shown in Fig. 28.

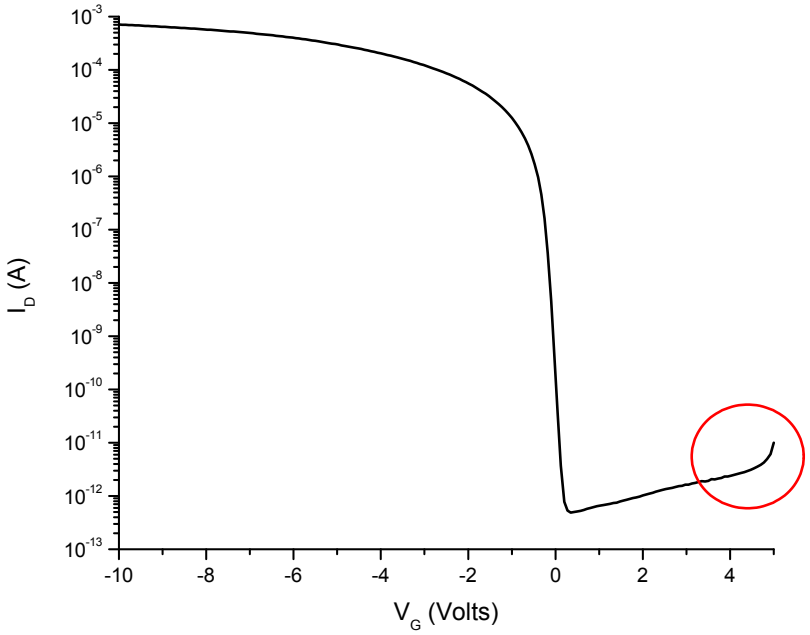

Fig. 28. Example I_D-V_G characteristics showing filling of trap states which suppresses the overall GIDL. V_{DS} = -5 V.

4.6 Stress Variability

This experiment involved the stressing of five different transistors of length = 12 µm and width = 24 µm and examining the how the ratio of post-stress to pre-stress drain current

varies. This experiment gave an estimate of both the uniformity of devices across the wafer along with some 3sigma error bars for data presented in section 6.2. The stress voltages chosen for this experiment were V_G = 15 V and V_D = -6 V giving an overall field of roughly 4.2 MV/cm over the gate oxide. The resulting standard deviation of the post-stress to pre-stress drain current ratio was 2.3% and the mean was 9.4% giving. Calculated three sigma limits would give ratios of anywhere between 2.5% and 15.3%. The results of this experiment show that once the field criteria for suppression is met there is a maximum value of suppression that is dominated by other mechanisms.

CHAPTER 5 – SCANNING PROBE MICROSCOPY

5.1 Introduction

A direct method used to investigate the mechanism of GIDL suppression was electrostatic force microscopy (EFM). The EFM technique is relatively new application of scanning probe microscopy involving a functionalized tip. The basic theory behind AFM, and EFM is described in the following subsections. EFM has demonstrated the ability to measure potential distributions and localized charge at the nanometer scale [19]. This technique was applied to investigate the induced charge responsible for GIDL suppression, providing a direct measurement rather than inference through the interpretation of device characteristics.

5.2 Basic Atomic Force Microscopy

Atomic force microscopy is a widely used metrology technique that has applications in several different industries. AFM based techniques fall into the broader field of scanning probe microscopy (SPM). An atomic force microscope is much like a scaled version of a stylus profilometer. The AFM has a micromachined cantilever with a sharp tip on the end of it. The tips are can be made of many different materials and have varying dimensions for a variety of different applications. Common tip materials include both silicon and silicon nitride. Tip radius is normally on the order of 1-10's of nanometers depending on how it is made and whether it is coated with a conductive material. A scanning electron micrograph of a sample cantilever and tip is shown in Fig. 29.

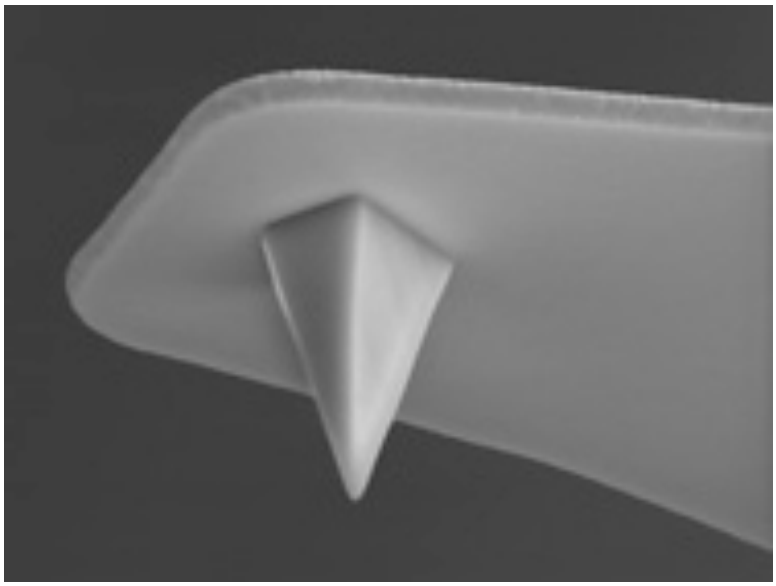

Fig. 29. SEM of typical AFM tip [20]

There are three basic methods of performing an AFM measurement: The tip can be dragged along the surface while in contact (contact mode), tapped along the surface (intermittent contact mode), or oscillated above the surface at a distance where there are considerable interactions between the tip and the substrate (non-contact mode). These modes of operation can be described using a simple two-particle interaction. The regions of operation are shown in Fig. 30 where a plot of force vs. tip-sample distance is shown.

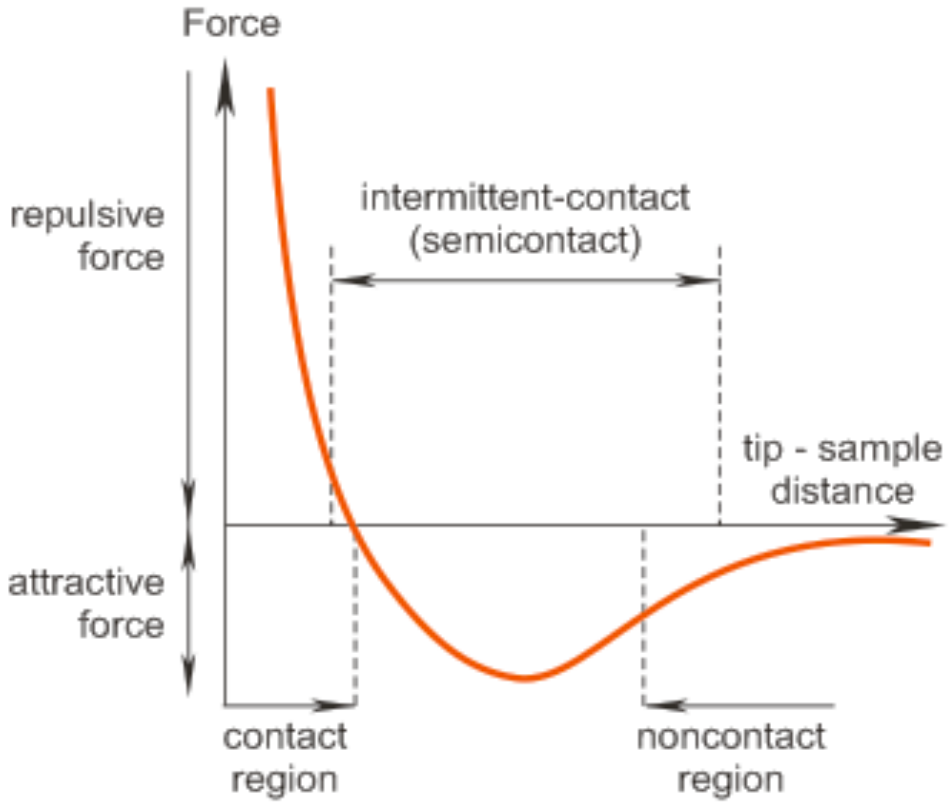

Fig. 30. Force vs. Tip-Sample Distance [21]

To monitor the deflection of the cantilever a laser is reflected off of the tip of the cantilever and up into a position sensitive photodetector. The position of laser on the detector correlates to the deflection of the cantilever. The detector consists of four quadrants of photodiodes where electron-hole pairs are generated when the laser illumination is incident upon them, thus providing the signal.

To create an image using an AFM the tip or sample needs to be moved in the X-Y plane. Once the scan size and scan frequency are set, the sample or tip is raster-scanned to form an image. To most common method to move the stage uses a piezoelectric tube scanner. The tube consists of four sections of a piezoelectric material. When a bias is applied to the material it will either expand or contract. If the opposite sides of the tube

have opposite biases, the tube will bend in a given direction. An image of both the laser based cantilever deflection detection system and a piezoelectric tube scanner is shown in Fig. 31.

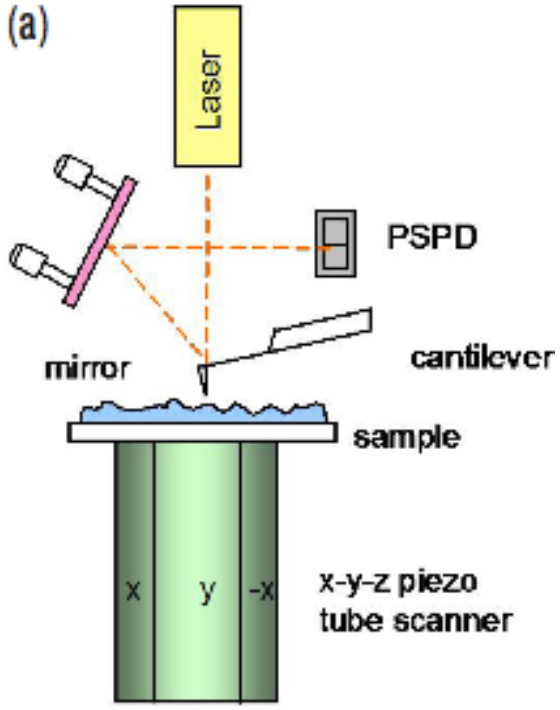

Fig. 31. Laser detector and piezoelectric tube scanner [22].

One of the major drawbacks of the piezoelectric tube scanner is crosstalk between the various scan directions. This means that the X, Y, or Z direction is changed the other two directions will also experience a change too. The crosstalk in the piezoelectric tube scanner can lead to curvature of the image and errors in measurement [22]. The AFM used for this study uses a different method to raster the sample that completely eliminates crosstalk between the X and Y axes. This system uses a flexure stage for the XY motion and a piezoelectric actuator for the Z motion. The system consists of a single module parallel-kinematics stage with low inertia and minimal out-of-plane motion [22]. To

incorporate an on-axis optical microscope, the laser system is also modified on the tool. A more accurate diagram of the laser detector is shown in Fig. 32.

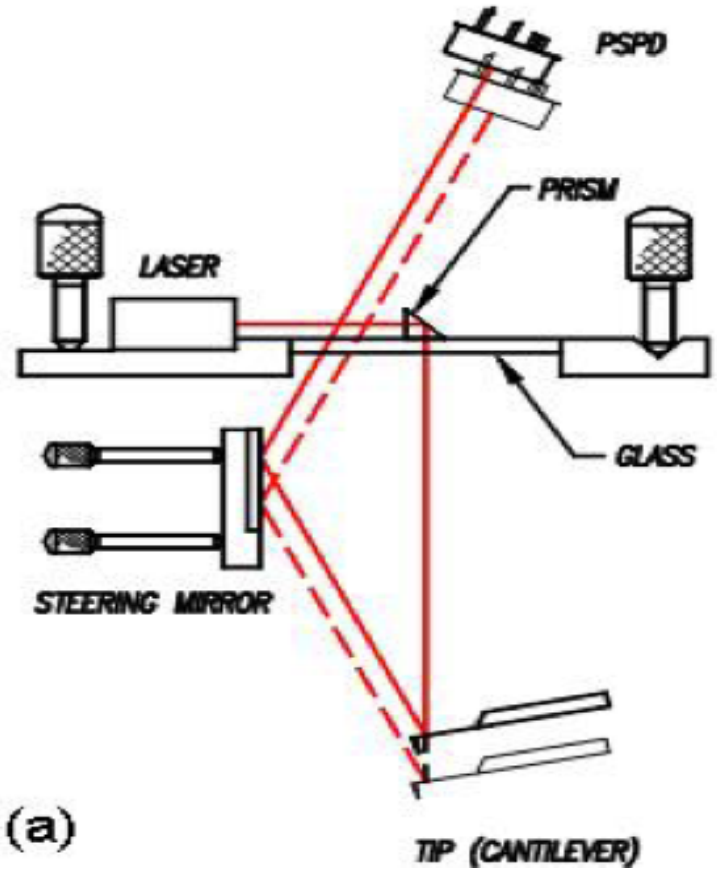

Fig. 32. Laser tip deflection detector system on PSIA XE-150 [22].

Contact Mode AFM

Contact mode AFM is exactly what the name implies; the tip is brought into intimate contact with the sample. The cantilever is lowered until it is within the range of the van der Waals forces acting on it from the surface. These forces are felt at distances a few 10s of angstroms away from the surface. When the tip keeps getting closer the repulsive force begins to act on it and the cantilever deflects. Note that a true

representation of van der Waals forces can only really be observed when the device is under ultra-high vacuum. If the device is in air, there will be some water adsorbed to the surface of the sample and the capillary force from the water will pull the tip into contact with the sample [21]. Once the tip is in contact there is enough force acting on the cantilever to cause a deflection.

There are two methods of scanning the surface of the sample in contact mode. The first method of scanning is constant height mode. Constant height mode is when the Z-height of the base of the cantilever is kept constant and the tip deflection is monitored to determine the topography. The other scanning method is called constant force imaging. Constant force imaging is when the Z-height of the base of the cantilever is adjusted using feedback circuitry to maintain a constant force (deflection) on the cantilever. The change in force on the cantilever is used as a feedback loop for the Z-height of the cantilever. The deflection from the upward force on the cantilever can be determined using Hooke's Law shown in (10) where k is the spring constant of the cantilever and x is the deflection of the cantilever.

$$F = -kx \tag{10}$$

Contact mode AFM has advantages and disadvantages. One of the main advantages is that contact mode AFM is fast relative to the other modes; the speed is limited only by the resonant frequency of the cantilever [21]. Although contact mode AFM is fast, the tip makes direct contact with the sample and could cause damage to both the tip and the

sample. The life of a contact AFM tip is much shorter than tips used in other modes of AFM. Sample damage may be a concern, especially if the sample was soft.

Intermittent Contact Mode AFM

Intermittent contact mode is commonly referred to as tapping mode AFM. In tapping mode the cantilever is oscillated at a frequency close to its own resonant frequency. The oscillation is created with a piezoelectric material. Since the oscillator is piezoelectric, the amplitude of oscillation is around 100nm. The tip is brought to a distance from the sample at which the van der Waals forces act on the oscillating tip and dampen the oscillations. Since the oscillation amplitude is so high the tip is in intermittent contact with the sample. Once the oscillations have been dampened, the sample is raster-scanned while keeping the amplitude of oscillation constant. The Z-Height of the cantilever is changed to keep the oscillation amplitude constant producing the topographical image. This method is similar to the constant force mode in contact AFM except now instead of measuring deflection the oscillation amplitude is measured and fed back into the system. Tapping mode has benefits over contact mode because the force applied to the sample is less than the force applied during contact mode. The tip life for intermittent contact mode AFM is much longer than that of contact mode AFM. This means that a broader range samples can be measured without the damage that would normally be induced from contact mode AFM.

Non-Contact Mode AFM

Non-contact mode AFM is similar to tapping mode AFM. The main difference is the oscillation amplitude of the cantilever. In tapping mode AFM the amplitude of oscillation is on the order of 100 nm while in non-contact mode AFM the amplitude of

oscillation is on the order of 1 nm. To achieve true non-contact mode AFM, the tip oscillation amplitude needs both accurate and precise. Once the tip is oscillating it is then lowered to a distance 50-100 angstroms from the surface so the tip begins to experience the van der Waals forces from the sample surface. The tip in this case is oscillated so that AC analysis techniques, such as feedback and amplification, can be used in the signal processing.

5.3 Electrostatic Force Microscopy

Electrostatic force microscopy (EFM) is a technique most commonly used in conjunction with non-contact mode AFM. The cantilever tip is coated with a conductive material that can be biased. EFM is typically a two-pass technique. The first pass uses conventional means to acquire the topography of the sample and then the second pass measures the electric field distribution underneath the tip. During the second pass a tip-bias is applied and the height of the cantilever is raised. The voltage between the tip and sample can be expressed using (11)[23] where V_{dc} is the dc bias of the tip V_s is the sample bias and V_{ac} and ω are the amplitude and frequency of the applied AC tip voltage.

$$V(t) = V_{dc} - V_s + V_{ac}sin(\omega t) \tag{11}$$

The reason for raising the cantilever height is so that the tip is in an electrostatic force dominated regime. Van der Waals forces decay proportional to $1/r^6$ while electrostatic forces decay as a function of $1/r^2$ so when the tip is lifted far enough away from the surface, electrostatic forces will dominate. When doing the electrostatic force scan, the Z-height of the cantilever also follows the topography (path of constant van der Waals

force) generated by the first scan. An illustration of the EFM scan technique is shown in Fig. 33.

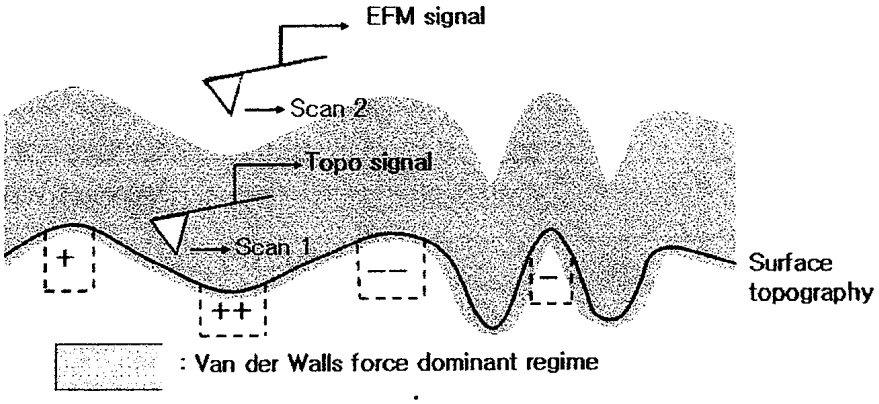

Figure 2 (a) Scanning in Standard EFM

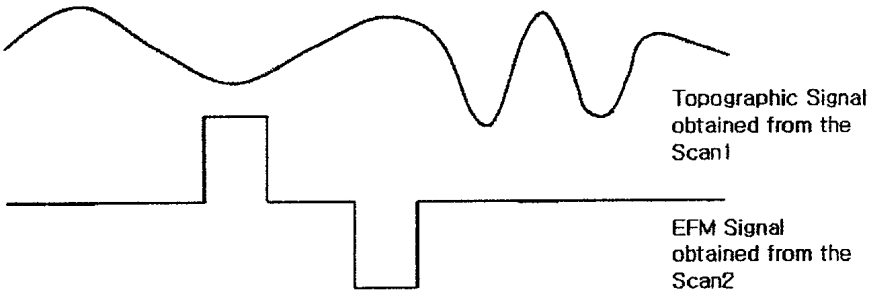

Figure 2 (b) Obtained Signals in Standard EFM

Fig. 33. EFM scanning technique [23].

There are also more enhanced modes of EFM that use external lock-in amplifiers to separate different signals. Assuming a parallel plate geometry and using (11) the force between the tip and sample can be calculated. The calculation of tip-sample interaction force with respect to time is shown in (12) [23].

$$
\begin{aligned}
F = qE = \frac{qV}{d} = \frac{CV^2}{d} \\
F(t) = \frac{C}{d}V(t)^2 \\
= \frac{C}{d}((V_{dc} - V_s)^2 + \frac{1}{2}V_{dc}^2) \\
+2(\frac{C}{d})(V_{dc} - V_s)V_{ac}sin(\omega t) \\
-\frac{C}{2d}V_{ac}^2 cos(2\omega t)
\end{aligned}
\tag{12}
$$

In the resulting solution there are three terms where *C* is the capacitance, *V* is electric potential, and *d* is the tip-to-sample distance. There is a DC term (no frequency dependence), an AC term with frequency of ω, and an AC term with frequency of 2ω. The first signal can be analyzed using the data acquisition software alone while the two AC components are sent through a lock-in amplifier that can read either part of the signal [23]. The DC part of the signal has contributions from both the capacitance and the surface potential. The AC signal with frequency ω contains both capacitance and surface potential as well, but the signal with frequency of 2ω only contains the capacitance component. If the ω is normalized with the 2ω signal the capacitance term can be eliminated thus solving for the surface potential of the sample.

5.4 Characterization of Charge in Insulators Using EFM

There have been multiple papers published on the use of an AFM/EFM system to characterize various insulators, such as the gate insulator of a transistor [19, 24]. A previous study examined the charge trapping and detrapping in a gate insulator that was grown as part of a 2μm localized oxidation of silicon (LOCOS) based process [19]. The gate oxides that were measured were not completely fabricated transistors; instead the authors used contact mode AFM with a functionalized tip to stress the sample with ±10V

for 60s and then measured it with various different tip-voltages in EFM mode. The EFM results from the stressed gate oxide are shown in Fig. 34.

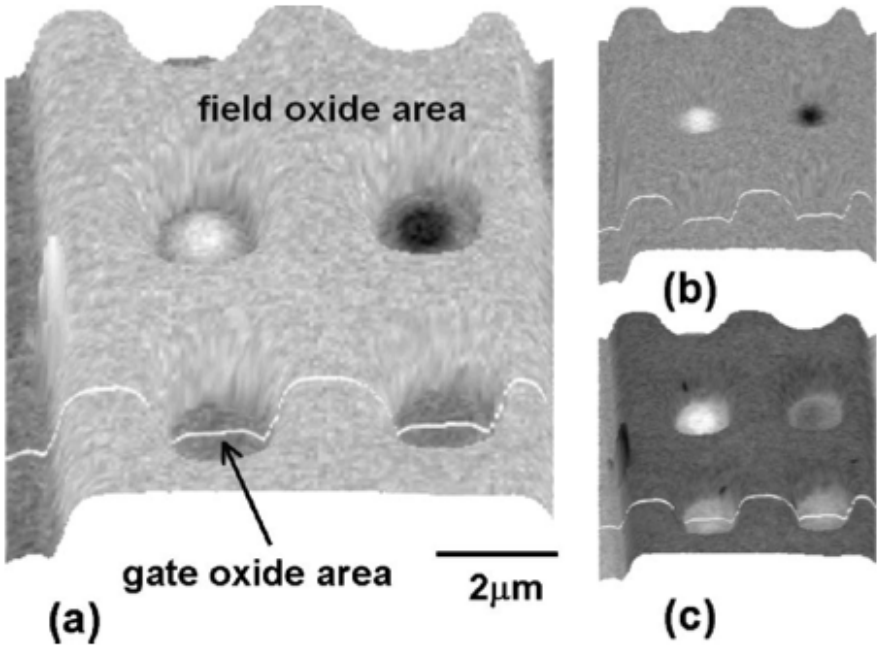

Fig. 34. EFM images at various tip-sample voltages (a) -2 V (b) 0 V (c) +2 V. The dark spots on each image indicate negative charge while the bright spots indicate positive charge. The white line shows the topography of the image (field to gate oxide) [19].

To convert the EFM signal into a localized charge concentration it needs to be converted into a contact potential signal. Once the contact potential is known it is compared to a known contact potential on the sample and then the charge can be extracted by using the model for a simple parallel plate capacitor [25]. The equation for this relationship is shown in (11) where σ_s is the charge underneath the tip, d is the thickness of the dielectric, and ε is the relative permittivity of the dielectric [25].

$$V_c(\sigma_s) = V_c + \frac{\sigma_s d}{\epsilon \epsilon_0} \tag{11}$$

5.5 EFM Sample Measurements

Transistor Deprocessing

Since any EFM based measurement needs to be performed without a gate metal on the gate-oxide. A technique has been developed to deprocess the transistors after the application of bias-stress. The deprocessing involves three main steps: etching of the aluminum, etching of the inter-layer dielectric (ILD), and etching of the molybdenum. In the deprocessing any plasma processes are avoided so that there is no charge unintentionally induced into the gate oxide. The aluminum etch is performed in a mixture of phosphoric, acetic, nitric acid, and water in a ratio of 16-1-1-2 respectively, heated to a temperature of 50 ºC. The etch rate of the aluminum etch is strongly dependant on temperature and 50 ºC is a fairly aggressive temperature. At 50 ºC the approximate etch rate is 600 nm/min. Although an etch rate is given, this etch is not timed. During the etch, the sample is examined and when the aluminum is gone the sample is kept in the etchant to make sure everything is clear then the sample is removed. The most important step of the deprocessing is the etching of the inter-level dielectric. This step is so important because it exposes the molybdenum gate metal for etching but also if there is an overetch, the gate oxide may be undercut and all the charge due to high field stress would be lost. The ILD etch is done in 10:1 buffered oxide etch. 10:1 buffered oxide etch contains a ratio of 10 parts of 40% ammonium fluoride, and 1 part 49% hydrofluoric acid. The etch rate is roughly 100 nm/min for LTO and 50 nm/min for thermally grown oxides. For removal of the molybdenum gate metal a PAN based etch is used. PAN etch has shown to be more effective in etching molybdenum at room temperature rather than the elevated temperatures used for etching of aluminum. The

etchant has also been seasoned with molybdenum, which also increases its effectiveness. With the molybdenum etch there is no concern about an endpoint because it is selective to oxide.

EFM Imaging

The objective of EFM measurements in this study was to be able to directly image the location and magnitude of charge in the gate oxide. To refine the EFM technique on the PSIA XE-150 a standard sample was used. This sample consisted of interdigitated gold fingers with a 10 μm pitch (line + space). The thickness of the gold fingers was 80nm. The interdigitated fingers allow one set of fingers to be grounded and the other set to have an applied bias. The topography image will show all of the fingers and the EFM image should show the fingers that have a bias on them. The amplitude of the EFM image is dependant on both the DC tip bias and the amplitude of the AC signal applied to the tip. The topography and EFM amplitude results from the measurement of the standard sample using enhanced EFM mode are shown in Fig. 35 and Fig. 36 respectively.

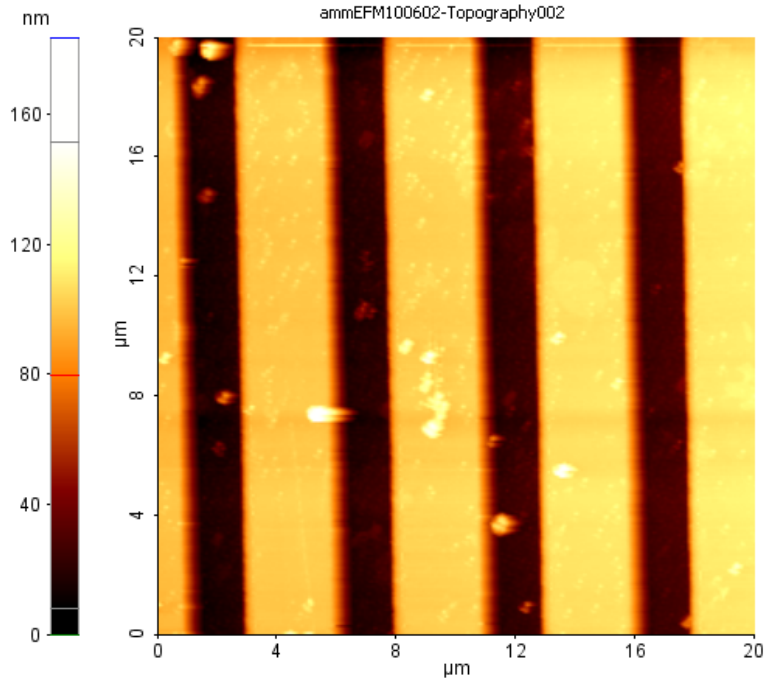

Fig. 35. Topography image standard EFM sample measured using enhanced EFM mode.

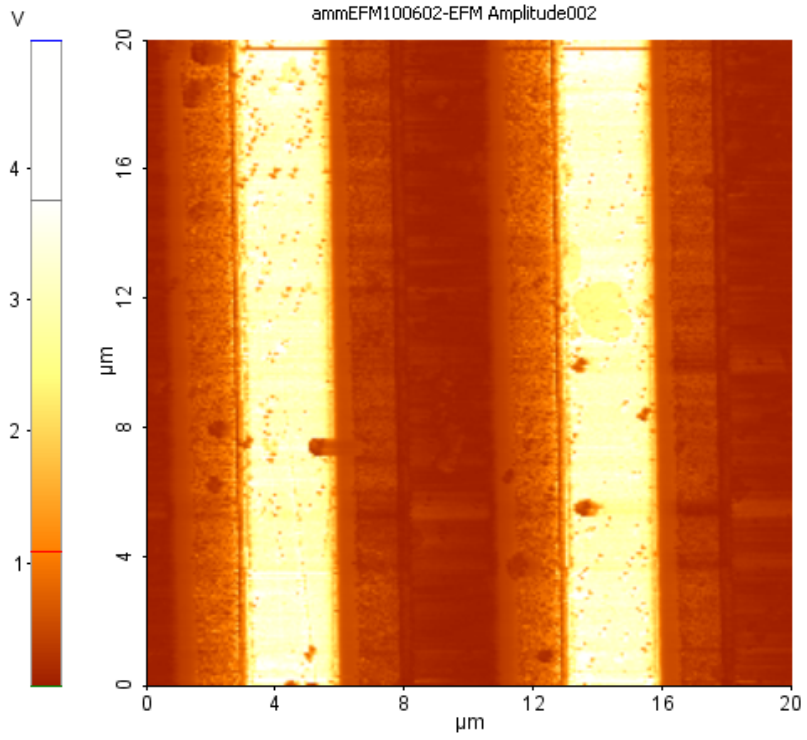

Fig. 36. EFM amplitude image of standard EFM sample using enhanced EFM mode. Tip bias = -2 V and Sample bias = 4 V

The amplitude image shows that only two of the four imaged fingers are showing a voltage. The voltage that is measured is dependant on the bias conditions. Another mode of EFM can be used to determine the exact potential of the sample. This mode is called scanning Kelvin probe microscopy (SKPM). SKPM uses tip bias as a feedback loop to minimize the force between the sample and the tip. When the force has been minimized the potential on the tip is assumed to be the potential on the sample. The method works well if there is no topography but if there is topography the SKPM signal can show irregularities due to the tip/sample interaction on an edge of a feature. An overlay of the topography and the EFM amplitude is shown in Fig. 37.

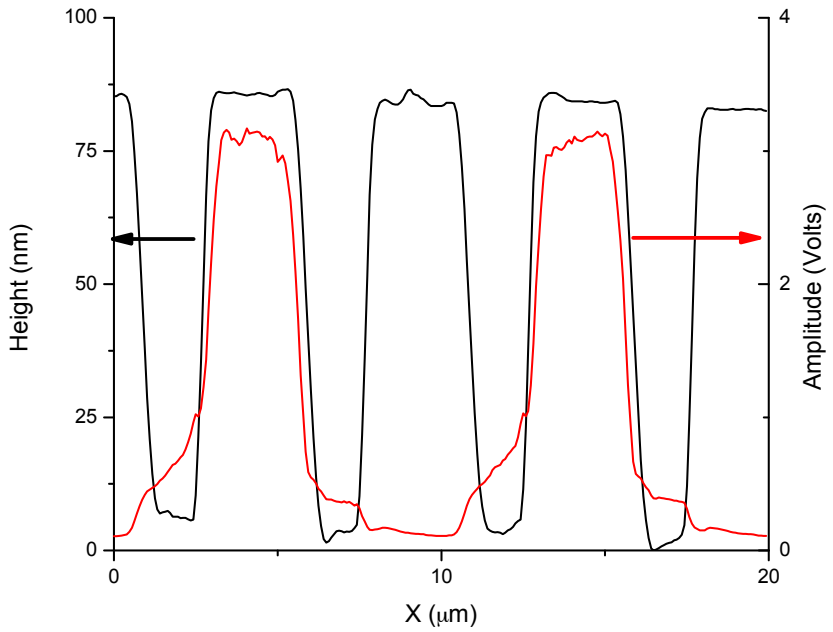

Fig. 37. Overlay of measured EFM topography and amplitude from standard EFM sample. Tip bias = -2 V and sample bias = 4 V.

The overlay confirms what is shown in the topography and amplitude images; every other finger shows a potential of approximately 3 V. In between the fingers there seems to be

some charging of the insulator, which appears to be a linear potential drop analogous to that seen in a parallel plate capacitor.

Once the technique was mastered on the standard sample, the deprocessed transistors were measured. In this experiment one 6 µm long by 24 µm wide transistor was stressed with a stress field equal to 4.2 MV/cm for 30 seconds and compared with a transistor of the same dimension that was not put under the same stress. The resulting images from the transistor that was stressed are shown in the following figures.

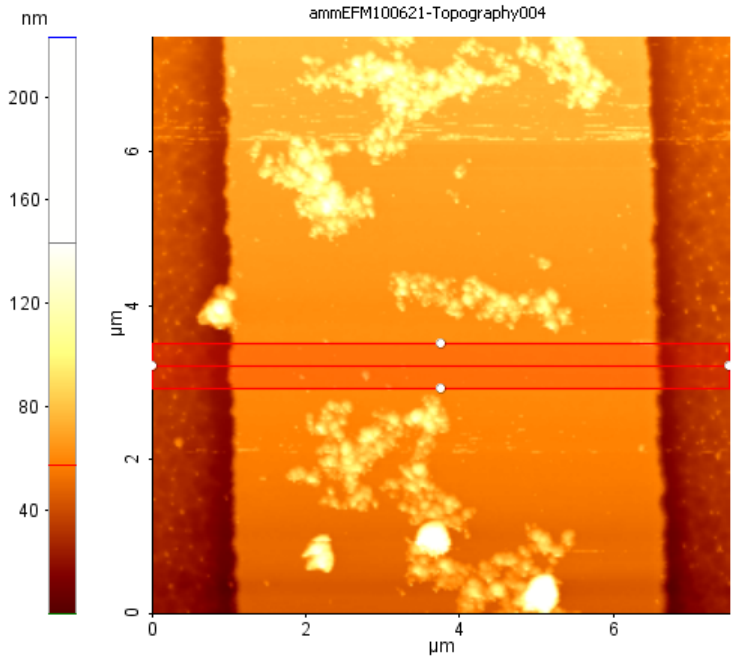

Fig. 38 Measurement of topography across the gate oxide of a deprocessed W/L = 24 µm / 6 µm transistor with a high-field stress of V_{DS} = -6 V and V_{GS} = 15 V for 30 seconds. Residue left on the gate oxide is most likely a precipitate that was not completely removed during the deprocessing.

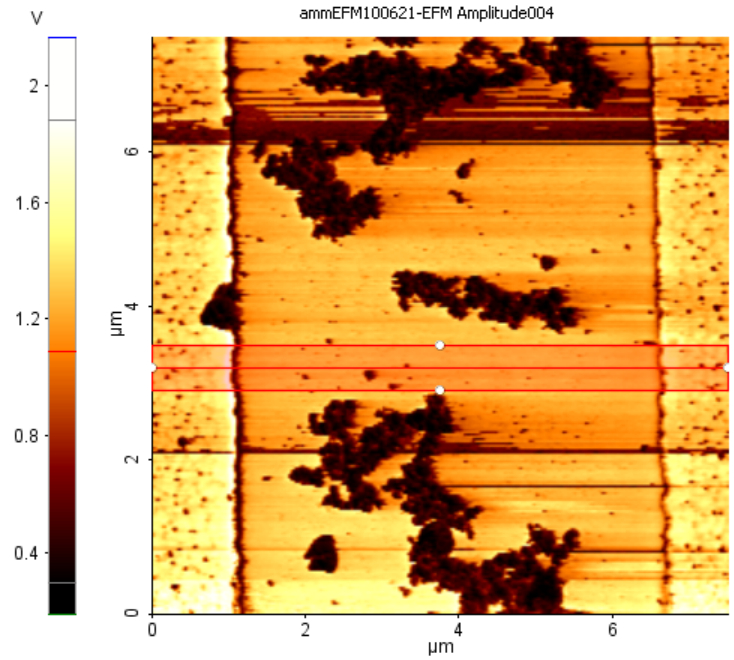

Fig. 39 Measurement of EFM amplitude across the gate oxide of a deprocessed W/L = 24 µm / 6 µm transistor with a high-field stress of V_{DS} = -6 V and V_{GS} = 15 V for 30 seconds. Tip bias = -4 V. There is no change between source and drain end of the device.

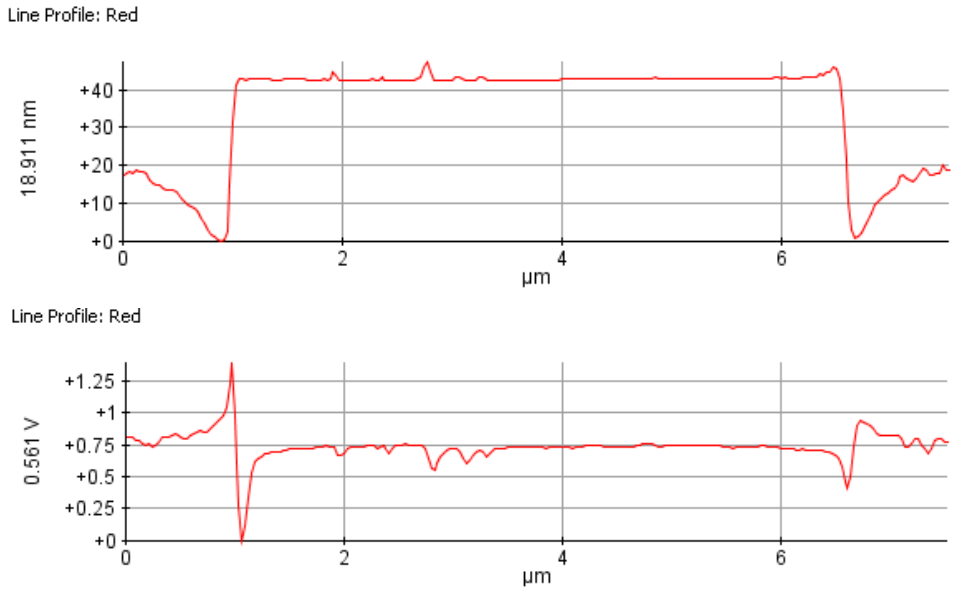

Fig. 40. Topography (top) and EFM amplitude (bottom) cut-lines of the device shown in Fig. 38 and Fig. 39 where the location of the cut-line is shown in red.

Fig. 38 and Fig. 40 show that the deprocessing of the transistor was successful. The cusp at around 1 µm and also at 7 µm on the cutline is thought to be due to the LTO deposition. Although there is some difference shown between the source and drain on the image/cutline of EFM amplitude, it is not certain that the difference is due to localized charge in the gate insulator or at the oxide/silicon interface. Comparison of the topography and EFM amplitude cutlines show that there is some correlation. To determine the amount of correlation, measurements were taken on a transistor that has had no high field stress. The results of the measurements are shown in the following figures.

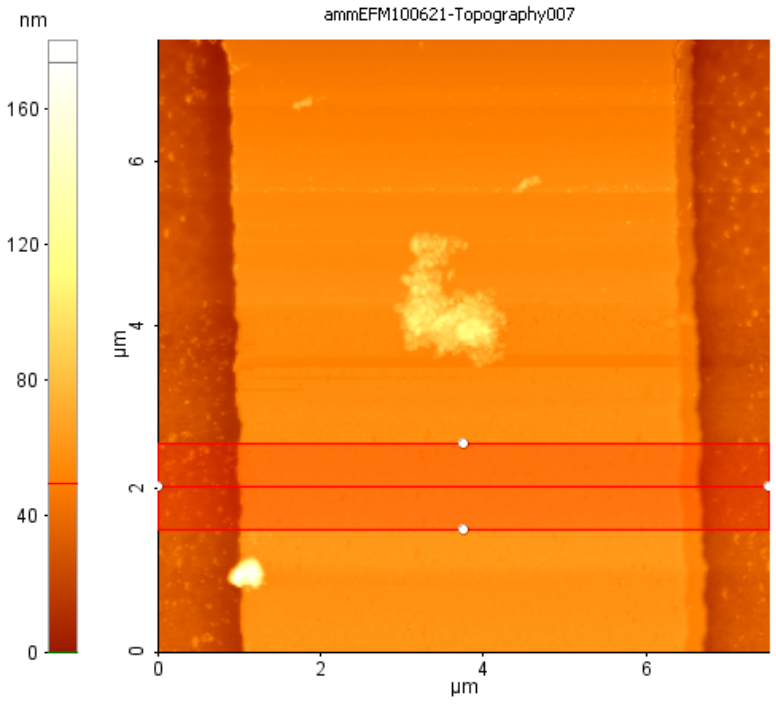

Fig. 41. Measurement of topography across the gate oxide of a deprocessed W/L = 24 µm / 6 µm transistor that has not been subjected to high-field stress. Residue left on the gate oxide is most likely a precipitate that was not completely removed during the deprocessing.

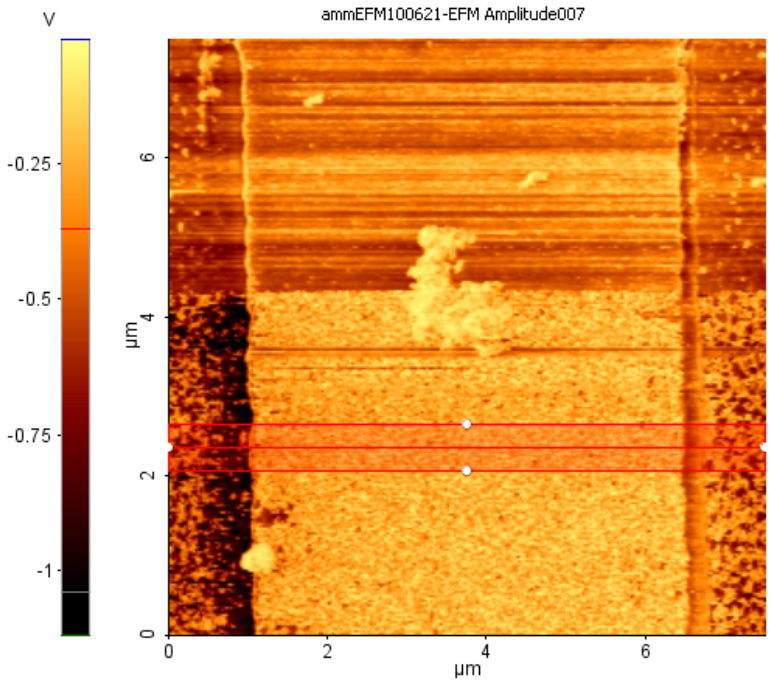

Fig. 42. Measurement of EFM amplitude across the gate oxide of a deprocessed W/L = 24 µm / 6 µm transistor. Tip bias = -4 V. There is no change between source and drain end of the device.

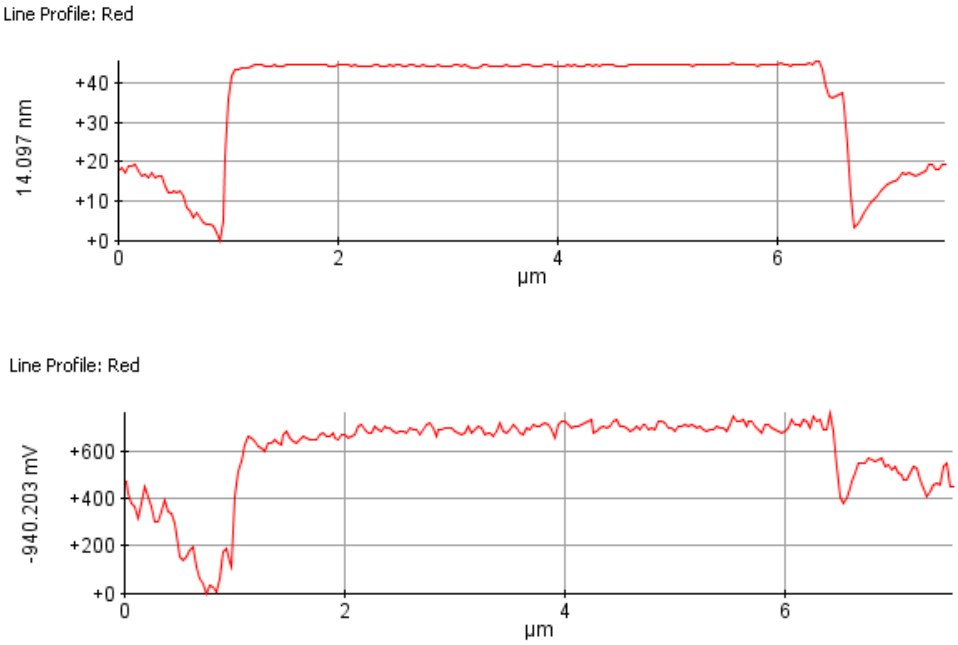

Fig. 43. Topography (top) and EFM amplitude (bottom) cut-lines of the device shown in Fig. 41 and Fig. 42 where the location of the cut-line is shown in red.

The previous figures show that there is no major difference in the EFM image of stressed and non-stressed devices. The cutlines in Fig. 43 show that there is also a difference between the drain and the source in the amplitude image, and that there is some correlation between amplitude and topography. The results from the EFM measurement show inconclusive evidence regarding the initial hypothesis that the mechanism of suppression is in-fact due to the creation and filling of interface traps near the drain-end of the transistor. The deprocessing of the transistors also could have influenced the amount of traps that were filled due to the higher temperature (50 ºC) etching of aluminum.

Although the measurements that were taken support what is seen in the electrical characterization, other attempts were made to measure different samples using the EFM. The first attempt was to compare the distribution of charge in oxide that had sputtered aluminum on top of it during a 425 ºC sinter with oxide that had the sputtered aluminum etched off of it before the sinter. Mercury probe capacitance-voltage (C-V) measurements were taken on the two treatment combinations and showed a 0.25 V shift in flat-band due to the sinter (passivation of charge). The shift that was seen in C-V showed no change in the EFM signal. The one thing that both the deprocessed transistor measurements and the sintered and non-sintered oxide measurements have in common are that both samples consist of silicon dioxide and the etching process involves the use of wet chemistry and water. SiO_2 is a hydrophilic material meaning that it likes to absorb and adsorb water. Water adsorbed to the surface can cause problems with and AFM technique. When lowering the tip into a region where van der Waals forces dominate there can also be capillary forces due to water adsorption to the surface of the sample.

Even when applying a bias to the tip, capillary forces can alter the AFM measurement at tip-sample distances of up to 300 nm [26]. The solution to the problem of water adsorption is the use of an AFM that is a temperature-controlled and pressure-controlled environment. If the measurement is taken in high-vacuum and the sample has been baked so that the water has desorbed the measurement will show no distortion due to capillary forces. Both references found that show measurements of charge concentrations in hydrophilic materials were taken in a high-vacuum (10^{-8} torr) environment [19, 24].

CHAPTER 6 – CONCLUSION

6.1 Introduction

This study explored the effects of high-field stress on the accumulation-mode PFET TFT. The effect has been both simulated and observed electrically. Attempts were also made to characterize the amount of charge using EFM methods. Results suggest that the effect is likely due to induced interface states at the drain edge of the TFT. This chapter presents an overview of the work that was completed and suggests future work that may be investigated in subsequent studies.

6.2 Simulation of the stress conditions

The device was simulated using Silvaco Atlas using the band-to-band tunneling model to show GIDL and the hot-electron model to show location of hot-carriers during the stress. Simulation showed results that were consistent with observations in the initial experimentation. The simulated device under stress showed that the hot-carrier region was within 200 nm distance from the drain/body junction. With an increase the level of negative charge in the hot-carrier region, the maximum electric field between the gate and drain decreases which in-turn decreases the band-to-band tunneling probability (GIDL). The results of the simulations showing the GIDL suppression are captured in Fig. 44 where complete suppression is shown for a device with $1x10^{12}$ cm^{-2} of interface charge at the drain end of the device.

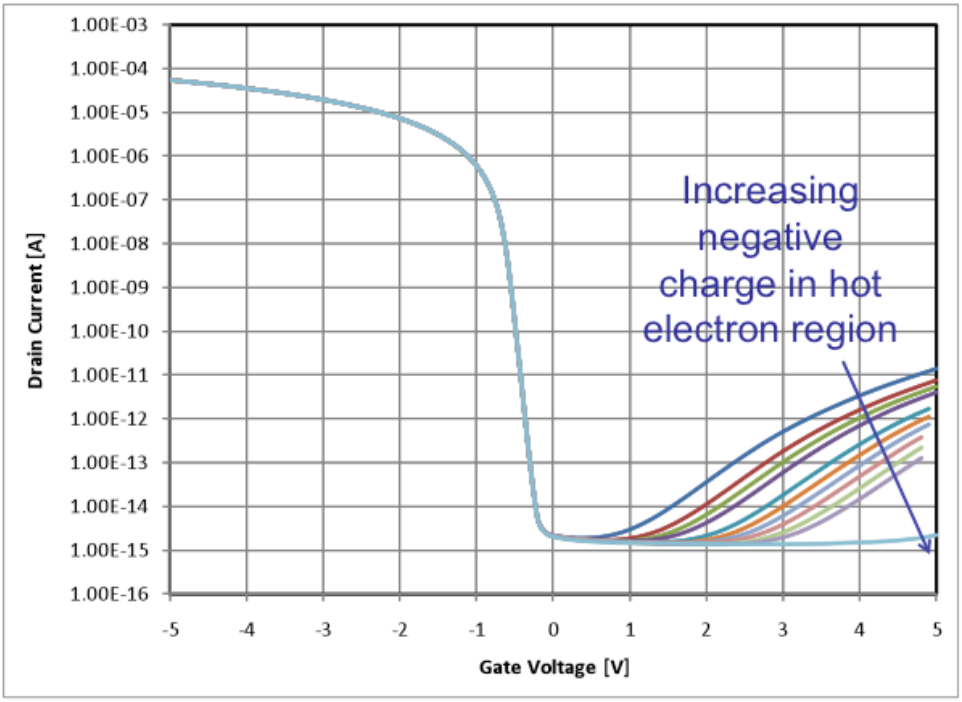

Fig. 44. Saturation-mode (V_{DS}=-5 V) subthreshold characteristics of various levels of charge starting with the virgin device then increasing charge level in increments of 10^{10} cm^{-2} up to 10^{11} cm^{-2}, with only the last curve showing 10^{12} cm^{-2} of charge. The last curve demonstrates suppression of GIDL at 5 V on the gate. Replicate of Fig. 14.

6.3 Fabrication and Electrical Characterization

This work presented successful fabrication of accumulation-mode PFETs on both SOI and SiOG substrates. I-V characteristics for the fabricated devices showed similar device characteristics as those that have been fabricated previously [1]. Electrical characterization showed that a normal stress field higher than ~4 MV/cm showed considerable GIDL suppression. The contour plot showing GIDL suppression with respect to stress voltages on the gate and drain is shown in Fig. 45, where the dark blue contour represents the maximum suppression.

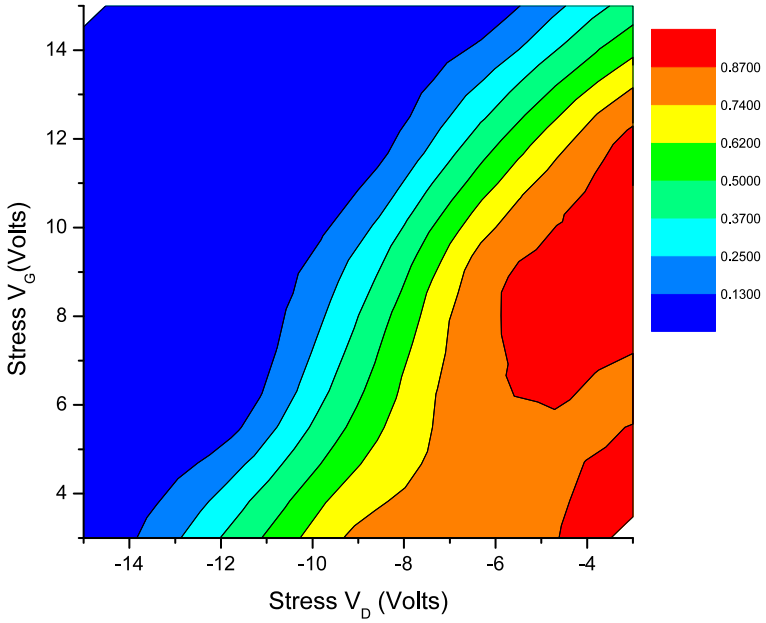

Fig. 45. Contour plot of I_D(post-stress)/I_D(pre-stress) measured at V_D=-10 V and V_G=5 V. Replicate of Fig. 23.

The effect of stress time was also studied and determined to have a significant impact on the post-stress device behavior. The transient study showed that GIDL suppression follows a time constant of ~15 sec, which supports the hypothesis that the mechanism of suppression is interface charge rather than fixed charge caused by the gate leakage current. The flux of electrons entering the gate oxide during the stress application was several orders of magnitude higher than the effective interface charge, which also suggests that the mechanism is not due to the charge injected into the gate oxide.

6.4 AFM Measurements

The measurements taken using AFM and EFM techniques showed mixed results. While the stressing and deprocessing of the accumulation-mode PFETs were successful,

there was no significant difference in EFM signal between the source and drain or devices measured with and without an applied stress. There are the following proposed reasons which may explain why there was no difference seen between devices that did and did not have an applied stress condition. Since it is proposed that the induced stress results in localized interface traps, the charge states may be in a neutral condition during the EFM measurement. In addition, the hydrophilicity of the SiO_2 could be affecting the measurement of the electrostatic potential that is imaged. All of the measurements were taken under humidity conditions of approximately 45%. Water molecules that adsorbs to the oxide surface can induce capillary forces on the AFM tip that can overwhelm other signals. A solution for water adsorption is for the AFM to be in a vacuum chamber with a hot chuck for a bake out, which is suggested for further investigation.

6.5 Final Remarks

The focus of this study was to investigate the effects of high-field induced stress suppression of GIDL in accumulation-mode PFETs. This work was accomplished through literature review, TCAD simulation, device fabrication, electrical characterization, and physical characterization. The TCAD model developed for the GIDL suppression was consistent with electrical results; the mechanism being a localized interface charge density close to the drain edge of the transistor. Electrical results also showed that this effect is both stress-field and stress-time dependant, and appears to be stable under DC conditions. The physical characterization of devices showed that a deprocessed gate oxide could be imaged. While in this specific instance EFM did not show definitive results, it is still believed that some SPM technique (e.g. scanning capacitance) can be used to quantify the location and density of the charge states created

from the induced stress under appropriate environmental conditions. Additional investigations are underway to better understand how other forms of stress influence the behavior of TFTs on both SiOG and SOI substrates.

APPENDIX A – ATLAS DECK

```
go atlas
mesh space.mult=2
set filename=draincharge
set fixed_charge=-1e12
x.mesh loc=0.00 spac=0.10
x.mesh loc=6.7 spac=0.005
x.mesh loc=7.3 spac=0.005
x.mesh loc=10 spac=0.10
y.mesh loc=-4 spac=0.25
y.mesh loc=-0.5 spac=0.1
y.mesh loc=0 spac=0.02
y.mesh loc=0.2 spac=0.02
y.mesh loc=0.25 spac=0.1

region number=1 x.min=0 x.max=30 y.min=-0.5 y.max=0.0 material=SiO2
region number=2 x.min=0 x.max=30 y.min=0 y.max=0.2 material=Silicon
region number=3 x.min=0 x.max=30 y.min=0.2 material=SiO2
region number=4 x.min=0 x.max=30 y.min=-4 y.max=-0.5 material=Silicon
electrode name=gate number=1 x.min=2.3 x.max=7.7 y.min=0.25 y.max=0.3
electrode name=drain number=2 x.min=10 x.max=10 y.min=0.0 y.max=0.2
electrode name=source number=3 x.min=0 x.max=0 y.min=0 y.max=0.2

contact name=gate work=4.53
contact name=source
contact name=drain
doping uniform conc=1e15 p.type regions=2
doping uniform conc=1e15 p.type regions=4

doping gaussian junction=1 conc=1e19 p.type x.right=3 lat.char=0.05 regions=2
doping gaussian junction=1 conc=1e19 p.type x.left=7 lat.char=0.05 regions=2
```

```
interface qf=1E11 x.min=0 x.max=7 y.min=0.2 y.max=0.2
interface qf=$fixed_charge x.min=7 x.max=7.2 y.min=0.2 y.max=0.2
interface qf=1E11 x.min=7 x.max=10 y.min=0.2 y.max=0.2
models srh conmob fldmob b.electrons=2 b.holes=1 evsatmod=0 hvsatmod=0 \
fermi bgn print numcar=2 temperature=300 \
impact hei fnord \
bbt.std bb.a=2e15 bb.gamma=2.0 bb.b=8e6
output hei hhi devdeg ox.charge con.band val.band u.bbt
method gummel newton carriers=2
#solve init
#solve vdrain=-0.1
#log outf=basedevicelin.log
#solve name=gate vgate=-5 vfinal=5 vstep=0.1
#log off
solve init
solve vdrain=-5
log outf=$'filename'_$'fixed_charge'.log
solve name=gate vgate=-5 vfinal=5 vstep=0.1
#solve name=gate vgate=0 vfinal=15 vstep=1
#solve name=drain vdrain=0 vfinal=-15 vstep=-0.5
log off
struct outfile=$'filename'_$'fixed_charge'.str
extract init infile="$'filename'_$'fixed_charge'.str"
#extract name="Em" max
extract name="Emax" 2d.max.conc impurity="Electric Field" material="Silicon"
#extract name="Emax X position" x.pos
#extract name="Emax Y position" y.pos
tonyplot
#tonyplot $"filename".log
quit
```

BIBLIOGRAPHY

[1] D. F. Dawson-Elli, C. A. K. Williams, J. G. Couillard, J. S. Cites, R. G. Manley, G. Fenger, and K. D. Hirschman, "Demonstration of High Performance TFTs on Silicon-on-Glass (SiOG) Substrate," *ECS Transactions,* vol. 8, pp. 223-228, 2007.

[2] SILVACO. (2007). *ATLAS User's Manual*.

[3] T. Y. Chan, J. Chen, P. K. Ko, and C. Hu, "The impact of gate-induced drain leakage current on MOSFET scaling," in *Electron Devices Meeting, 1987 International*, 1987, pp. 718-721.

[4] J. Chen, T. Y. Chan, I. C. Chen, P. K. Ko, and C. Hu, "Subbreakdown drain leakage current in MOSFET," *Electron Device Letters, IEEE,* vol. 8, pp. 515-517, 1987.

[5] K. Kurimoto, Y. Odake, and S. Odanaka, "Drain leakage current characteristics due to the band-to-band tunneling in LDD MOS devices," in *Electron Devices Meeting, 1989 International*, 1989, pp. 621-624.

[6] J. Chen, F. Assaderaghi, P. K. Ko, and C. Hu, "The enhancement of gate-induced-drain-leakage (GIDL) current in short-channel SOI MOSFET and its application in measuring lateral bipolar current gain β " *Electron Device Letters, IEEE,* vol. 13, pp. 572-574, 1992.

[7] S. M. Sze and K. K. Ng, *Physics of Semiconductor Devices*, Third Edition ed.: John Wiley & Sons, Inc., 2007.

[8] J. E. Moon, T. Garfinkel, J. Chung, M. Wong, P. K. Ko, and C. Hu, "A new LDD structure: total overlap with polysilicon spacer (TOPS)," *Electron Device Letters, IEEE,* vol. 11, pp. 221-223, 1990.

[9] S. A. Parke, J. E. Moon, H. J. C. Wann, P. K. Ko, and C. Hu, "Design for Suppression of Gate-Induced Drain Leakage in LDD MOSFET's Using a Quasi-Two-Dimensional Analytical Model," *IEEE Transactions on Electron Devices,* vol. 39ß, pp. 1694-1703, Jul 1992.

[10] B. Zhang, A. Balasinski, and T. P. Ma, "Hot-carrier effects on gate-induced-drain-leakage (GIDL) current in thin-film SOI/NMOSFET's," *IEEE Electron Device Letters,* vol. 15, pp. 169-171, 1994.

[11] T. Wang, T. E. Chang, L. P. Chiang, C. H. Wang, N. K. Zous, and C. Huang, "Investigation of oxide charge trapping and detrapping in a MOSFET by using a GIDL current technique," *IEEE Transactions on Electron Devices,* vol. 45, pp. 1511-1517, 1998.

[12] I. C. Chen, J. Y. Choi, T. Y. Chan, T. C. Ong, and C. Hu, "The effect of channel hot carrier stressing on gate oxide integrity in MOSFET," in *26th Annual Proceedings. Reliability Physics 1988 (Cat. No.88CH2508-0), 12-14 April 1988*, New York, NY, USA, 1988, pp. 1-7.

[13] R. G. Manley, "Operation of an Accumulation Mode PFET," ed. Rochester, 2010.

[14] G. K. Celler and S. Cristoloveanu, "Frontiers of silicon-on-insulator," *Journal of Applied Physics,* vol. 93, p. 4955, 2003.

[15] E. M. Woodard, R. G. Manley, G. Fenger, R. L. Saxer, K. D. Hirschman, D. Dawson-Elli, and J. G. Couillard, "Low Temperature Dopant Activation for

Integrated Electronics Applications," in *University/Government/Industry Microelectronics Symposium, 2006 16th Biennial*, 2006, pp. 161-168.

[16] R. G. Manley, "Development and modeling of a low temperature thin-film CMOS on glass," MS, Microelectronic Engineering, Rochester Institute of Technology, Rochester, 2009.

[17] C. Hu, C. T. Simon, H. Fu-Chieh, K. Ping-Keung, C. Tung-Yi, and K. W. Terrill, "Hot-electron-induced MOSFET degradation--Model, monitor, and improvement," *Electron Devices, IEEE Transactions on,* vol. 32, pp. 375-385, 1985.

[18] R. Manley, G. Fenger, P. Meller, K. Hirschman, C. A. K. Williams, D. Dawson-Elli, J. G. Couillard, and J. Cites, "Development of Integrated Electronics on Silicon-on-Glass (SiOG) Substrate," *ECS Transactions,* vol. 16, pp. 371-380, 2008.

[19] G. H. Buh, H. J. Chung, and Y. Kuk, "Real-time evolution of trapped charge in a SiO[sub 2] layer: An electrostatic force microscopy study," *Applied Physics Letters,* vol. 79, 2001.

[20] (2009, November 11). *Veeco - MSCT-EXMT-BF1*. Available: http://www.veecoprobes.com/p-3423-msct-exmt-bf1.aspx

[21] (2009, November 11). *SPM Principles*. Available: http://www.ntmdt.com/spm-principles

[22] (2009, November 11). *XE Mode Note: Technology Behind the XE-series AFM*. Available: http://www.parkafm.com/New_html/resources/02mode.php

[23] *Operation Manual for the EFM*. Santa Clara, CA: Park Systems.

[24] S. D. Tzeng and S. Gwo, "Charge trapping properties at silicon nitride/silicon oxide interface studied by variable-temperature electrostatic force microscopy," *Journal of Applied Physics,* vol. 100, p. 023711, 2006.

[25] M. Saint Jean, S. Hudlet, C. Guthmann, and J. Berger, "Charge dynamics and time evolution of contact potential studied by atomic force microscopy," *Physical Review B,* vol. 56, p. 15391, 1997.

[26] K. Yamamoto, C. Tanuma, and N. Gemma, "Competition between electrostatic and capillary forces acting on a single particle," *Japanese Journal of Applied Physics, Part 1 (Regular Papers & Short Notes),* vol. 34, pp. 4176-84, 1995.