INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps.
Nanometer Scale Electrical Characterization of Thin Dielectric Films

DISSERTATION

Presented in Partial Fulfillment of the Requirements for
the Degree Doctor of Philosophy in the
Graduate School of The Ohio State University

By

David T. Lee. B.A.. M.S.

* * * * *

The Ohio State University

2002

Dissertation Committee:

Prof. Jonathan Pelz, Adviser
Prof. Bharat Bhushan, Co-Adviser
Prof. Leonard J. Brillson
Prof. Charles A. Ebner
Prof. Linn Van Woerkom

Approved by

Co-Adviser

Co-Adviser

Physics Department
© Copyright by
David T. Lee
2002
ABSTRACT

This work is directed towards the use of electrical properties to characterize thin dielectric films on nm length scales. In particular, two technologically important systems have been studied: interface defects at the Si/SiO₂ interface and the use of scanning capacitance microscopy to investigate lubricant films, primarily composed of fully bonded perfluoropolyethers, that are used to lubricate hard disk drive platters and show promise for use in micro-electromechanical systems (MEMS).

The first system is the charge trapping defect found at the interface between Si and thin silicon dioxide films grown on the Si. The goal of this work is to make both ballistic electron emission microscopy (BEEM) and charge pumping measurements on the same device. This combination of techniques will allow us to make nm-scale measurements of interface state formation and hot-carrier transport within working metal oxide semiconductor field effect transistors (MOSFET). We have shown that BEEM measurements can be made on metal-oxide-semiconductor (MOS) capacitors that have been subjected to standard semiconductor fabrication processes. While BEEM compatible MOSFETs have not yet been produced, an ongoing effort in collaboration with IMEC in Leuven, Belgium is progressing towards working, BEEM compatible MOSFETs.

The second system under study is the use of capacitance measurements to resolve sub-nm variations in the thickness of thin dielectric films with nm-scale lateral
resolution. Towards this goal, we have: developed direct, low-frequency scanning capacitance microscopy (SCM) instrumentation capable of measuring $10^{-18}$ F (aF) changes in the capacitance between an atomic force microscope (AFM) tip and a sample with a noise level of $0.1$ aF/√Hz; for the first time, quantified and developed means of accounting for changes in parasitic capacitance that occur while scanning an AFM tip; for the first time, quantified the effective area of the meniscus that forms between the AFM tip and the sample while scanning in air; and made the highest lateral resolution (~ 200 nm) measurements of dielectric film thickness variations with ~ 1nm resolution.
I dedicate this work to my wife, Lyn Jakeman, and son, Max Lee, whose support and love make mine a gifted life.
ACKNOWLEDGMENTS

I wish to acknowledge the support and guidance of my mentor, Dr. Jonathan Pelz. He has encouraged me in my struggle through difficult and challenging times and pointed out my successes when they were hidden. I look forward to continuing to work with him in the future.

I would also like to acknowledge the guidance of Dr. James Stith, my first year academic advisor, without whom there would not have been a second year of graduate school. Along these lines, the extra effort of Dr. Eric Bratten to keep his door open and provide tutoring throughout my first year is also much appreciated.

Much practical help and guidance was also provided by Dr. George Valco, Mr. Tom Kelch, Ms. Rita Rokhlin, and the physics electronics shop, especially Mark Johnson. Of course, the physics department would cease to function if not for the efforts of the computer support group. Special thanks to Brian Dunlap, Brian Keller, and J.D. Wear.

The many contributions to this work by my fellow basement lab rats are also much appreciated. Thanks to Ben Kaczer, Jai Im, Jon-Fredrik Nielsen, Eric Heller, Yi Ding, Cristian Tivares, Brian Gibbons, and Kibog Park for all the conversations, lunches, and advice.

Last, but not least, I would like to thank my committee members for their advice, guidance, and patience with my last minute work habits!
The financial support for this report was provided by the National Science Foundation (Contract NOS. CMS 0001382 and DMR-0074416). The content of this information does not necessarily reflect the position or policy of the Government and no official endorsement should be inferred. Additional support was provided in the form of a seed grant from the Stanford Nanofabrication Facility.
VITA

31 March 1961
Born. Chicago, IL, USA

1992–1995
San Jose State University, San Jose, CA
Master of Science. Computational Physics.
Dr. Alejandro Garcia. Advisor

1978–1982
Maryville College, Maryville, TN
Bachelor of Arts. Physics and Chemistry Major.
Graduated Cum Laude

1995–1996
Columbus, OH

1992–1995
Programmer/Analyst. SETI Institute.
Mountain View, CA

1990–1992
First Officer, AMR Eagle/Wings West Airlines.
DFW, TX

1987–1990
Charter Pilot. Kenn Air Aviation.
Gainesville, FL

1986–1990
Flight Instructor. GatorAire.
Gainesville, FL

PUBLICATIONS


FIELDS OF STUDY

Major Field: Physics
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abstract ........................................................................................................ii</td>
</tr>
<tr>
<td>Dedication ........................................................................................................ iv</td>
</tr>
<tr>
<td>Acknowledgments .............................................................................................. v</td>
</tr>
<tr>
<td>Vita .................................................................................................................... vii</td>
</tr>
<tr>
<td>List of Tables ................................................................................................... xi</td>
</tr>
<tr>
<td>List of Figures .................................................................................................. xii</td>
</tr>
<tr>
<td>Chapters:</td>
</tr>
<tr>
<td>1. Overview ....................................................................................................... 1</td>
</tr>
<tr>
<td>2. Silicon Dioxide Interface States ............................................................ 6</td>
</tr>
<tr>
<td>2.1 Physical properties of thin SiO$_2$ films on Si ................................... 9</td>
</tr>
<tr>
<td>2.1.1 Geometry of Interface Structures ...................................................... 11</td>
</tr>
<tr>
<td>2.1.2 Bonding and Stoichiometry at the Interface ......................................... 13</td>
</tr>
<tr>
<td>2.1.3 Defects and the Nature of Interface States ........................................ 16</td>
</tr>
<tr>
<td>2.2 Electrical Characteristics of SiO$_2$ on Si .......................................... 19</td>
</tr>
<tr>
<td>2.2.1 Metal-oxide–semiconductor capacitor physics ...................................... 20</td>
</tr>
<tr>
<td>2.2.2 Ballistic Electron Emission Microscopy ............................................... 27</td>
</tr>
<tr>
<td>2.2.3 MOSFET Operation .................................................................................... 29</td>
</tr>
<tr>
<td>2.2.4 Charge pumping physics ........................................................................ 34</td>
</tr>
<tr>
<td>2.2.5 Detailed Charge Pumping Model ............................................................. 39</td>
</tr>
<tr>
<td>2.2.6 Interface Studies with Charge Pumping ............................................... 42</td>
</tr>
<tr>
<td>2.3 BEEM compatible MOSFET Fabrication ..................................................... 48</td>
</tr>
</tbody>
</table>

viii
## LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1 Implant Parameters. The source/drain implant was done by Nu Ions of San Jose, CA. All others were done at the SNF.</td>
<td>71</td>
</tr>
<tr>
<td>A.1 OSU initial wafer degrease. (RT is room temperature.)</td>
<td>131</td>
</tr>
<tr>
<td>A.2 OSU preoxidation cleaning recipes. This recipe is a variant of the industry standard “RCA clean”. SC1 is standard clean 1. SC2 is standard clean 2. and RT is room temperature. The HF dips marked with * are not used on gate oxides.</td>
<td>132</td>
</tr>
<tr>
<td>A.3 OSU photoresist strip and clean. (RT is room temperature.) Note that the ultrasonic baths are not used when there is a thin, metal film on the wafers.</td>
<td>133</td>
</tr>
<tr>
<td>A.4 OSU oxidation recipes: wet oxidations were done by bubbling O₂ through boiling deionized water.</td>
<td>134</td>
</tr>
<tr>
<td>B.1 Standard Cleaning Procedures. HF dips marked with * are not used on unprotected gate oxides.</td>
<td>138</td>
</tr>
<tr>
<td>B.2 Furnace Parameters for LOCOS steps. DCS = dichlorosilane</td>
<td>139</td>
</tr>
<tr>
<td>B.3 Gate oxidation growth conditions</td>
<td>140</td>
</tr>
<tr>
<td>C.1 The component values used in phase shift circuitry are listed here. The component numbers refer to Figure C.1.</td>
<td>143</td>
</tr>
</tbody>
</table>
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>Picture of an electrostatic motor produced with MEMS technology. Nanometer scale defects in lubricant films cause high stiction, friction, and early failure of these devices.</td>
</tr>
<tr>
<td>2.1</td>
<td>Schematic view of a MOSFET. Two reverse-biased diodes, the source (i) and drain (ii), are separated by a channel (iii). The conductance of this channel is modulated by the potential applied to the gate (iv). The gate oxide (v) is an insulating barrier that prevents current from flowing from the gate to the substrate.</td>
</tr>
<tr>
<td>2.2</td>
<td>Schematic view of (a) MOS capacitor components, (b) the band diagram for the separate components, (c) the assembled MOS capacitor, and (d) the band diagram for the MOS capacitor with the gate connected to the substrate.</td>
</tr>
<tr>
<td>2.3</td>
<td>Band bending diagrams for (a) accumulation, (b) flatband, (c) depletion, and (d) inversion for a p-type Si MOS capacitor.</td>
</tr>
<tr>
<td>2.4</td>
<td>(a) A schematic of a MOSCAP is a depiction of the applied voltages and equivalent capacitances for low frequency capacitance measurements. (b) A schematic low frequency CV characteristic for a MOSCAP showing the (i) accumulation, (ii) depletion, (iii) inversion onset, and (iv) deep inversion regimes. The blue line is a MOSCAP without interface traps; the red line depicts the broadening and shallowing of the CV characteristic as described in the text.</td>
</tr>
</tbody>
</table>
2.5 (a) A schematic of a MOSCAP is a depiction of the applied voltages and equivalent capacitances for high frequency capacitance measurements.
(b) A schematic CV characteristic for a high frequency MOSCAP showing the (i) accumulation, (ii) depletion, and (iii) inversion regimes. The blue line is a MOSCAP without interface traps: the red line depicts the broadening and shallowing of the CV characteristic as described in the text. 27

2.6 (a) A schematic of the band structure during a BEEM measurement. The STM tip bias $V_{tm}$ controls the energy of the tunneling electrons while the sample bias $V_{m}$ controls the voltage drop across the oxide barrier. 28

2.7 The band structure of the SiO$_2$/Si interface during MOSFET operation as a function of position is presented. In (a), $V_{dd} = 0$ and $V_{g}$ is at flat band voltage. In (b), $V_{dd} = 0$ and $V_{g}$ is above threshold voltage. In (c), $V_{dd} > 0$ and $V_{g}$ is above threshold voltage. 31

2.8 A schematic of diagram of the experimental setup for charge pumping measurements. 36

2.9 Schematic view of MOSCAP fabrication with a "lift-off" metallization process. (a) Photoresist is deposited on the SiO$_2$ layer and the top of the photoresist is hardened. (b) The photoresist is exposed, developed, and metal is evaporated onto the surface. (c) The photoresist is removed, taking the excess metal with it, and a MOSCAP remains. 51

2.10 Low-frequency $CV$ curve obtained from MOSCAP fabricated with a lift-off metallization. The metal square is 250 x 250$\mu$m. The AC frequency was 100 Hz. 53

2.11 Schematic of the experimental setup used to acquire BEEM images. The tip current $I_t$ is kept constant by the feedback system that controls tip height above the surface. The tip voltage $V_t$ is under computer control and the BEEM current $I_e$ is digitized by and stored on a computer. 54

2.12 Topography and BEEM images of a MOSCAP. The vertical scales are 4.5 nm for the topography image and 3 pA for the BEEM image. (data file: 9809236). 55
2.13 BEEM IV curves obtained at 6 different locations in a 150 x 150 nm region are shown. The color coding groups IV curves with similar threshold voltages. All of the IV curves shown have threshold voltages greater than the ~ 4V threshold voltage for SiO₂. (data file: 9809249).

2.14 (a) A scale drawing of the charge pumping device shows a top view. (b) A cross section along the dotted line, in (a), illustrates the topography of the device. (c) The gate oxide region is emphasized.

2.15 Schematic cross-sections through a MOSFET following: (a) field oxidation. (b) diffusion. (c) gate area exposure and gate oxidation. (d) contact via etch and contact metallization. and (e) gate metallization.

2.16 (a) BEEM compatible MOSFET fabrication process flow chart.

2.17 IV curves for three pairs of substrate contacts. Pairs 1-2 are much closer to each other than the other two pairs. The substrate resistance accounts for nearly all of the measured resistance.

2.18 Forward (a) and reverse bias (b) IV curves for the pn-junction formed by the charge pumping device diffusion region and the substrate. Note the non-ideal behavior (ideality factor ~ 2.5) and the high reverse bias leakage current. See the text for a discussion of the causes of failure in these diodes.

2.19 High (red line. 10 kHz) and low frequency (blue line. 100 Hz) CV curves for an OSU MOSCAP (W6AC2).

2.20 (a) An OSU charge pumping device, with the Al diffusion contacts removed by etching, is shown. (b) A 10x closer view of the edge of the contact via clearly shows the Al spikes that short the diffusion to the substrate.

2.21 Schematic view of the MOSFET fabrication process used at the SNF. (a) and (b) show views of a MOSFET during and following LOCOS. (c) shows the gate oxide and source/drain As implant. (d) shows the source/drain contact formation and (e) shows the completed MOSFET.

2.22 Flow diagram for the semi-recessed local oxidation of Si process used at the SNF.
2.23 Flow diagram for gate oxidation, source/drain implants, backside gettering implant, and dopant activation anneal ......................................... 72

2.24 Flow diagram for the source/drain contact formation and gate contact lift-off performed at the SNF. .......................................................... 74

2.25 Picture of a completed MOSFET. The probe damage was inflicted during electrical testing of this device. The vertical alignment error between contact metal and the active region was 4 μm; the horizontal alignment error was ~0.5 μm. ......................................................... 75

2.26 An IV characteristic between two substrate contacts shows low resistance (263 Ω) and excellent linearity. The dots are data points and the red line is a linear least squares fit to the data (Pearson's r = 0.999). ............................................. 76

2.27 A forward bias IV characteristic for an SNF diode shows good performance. The maximum reverse bias current observed (at -4 V) was 8 pA ........................................................................ 77

3.1 (a) schematic of the probe assembly-sample region (b) equivalent electric circuit includes nm-scale tip-sample capacitance $C_{ts}$ in parallel with the stray capacitance $C_{stray}$ and film resistance $R_{film}$. A conducting substrate is assumed. $C_{stray}$ is typically about 1/2 pF while $C_{ts}$ is typically expected to be, depending on tip radius, 1-100 aF. 82

3.2 (a) Schematic of a typical RCA type tuned circuit used in commercial SCMs. A UHF signal is inductively coupled to a tuned circuit that includes the AFM tip-sample capacitance. The tuned circuit is inductively coupled to a diode demodulator detector circuit. (b) Schematic of a bridge circuit for detecting small changes in capacitance against a large background capacitance $C_p$. The nulling capacitor is varied until there is no output voltage. Variations in $C_{ts}$ then unbalance the bridge and create an output voltage. A lock-in amplifier is typically used to amplify the signal and reject the noise. ........................................ 83

3.3 Schematic of circuitry used to measure capacitance. The left side includes the probe-sample region while the right side includes the “buckling circuit” described in the text. 85
3.4 (a) Noise spectra obtained with circuit connected directly to a fixed capacitor test circuit (solid line) and with ~30 cm coaxial cable inserted between capacitor and current amplifier (dashed line). (b) Noise spectra obtained with measurement circuit connected to the tip. The tip was on the sample surface and not scanning (solid line) and on the surface during a 10μm scan (dashed line). The non-scanning noise level is < 0.4 aF, which is similar to that measured with the test circuit. .......................... 88

3.5 (a) Topography image of a 104nm trench etched into SiO₂ on Si and covered with 10 nm of Si₃N₄. (b) Cross section of the topography along the dashed line in (a). (c) Simultaneously acquired capacitance image of the trench. (d) Capacitance cross section along the dashed line in (b). 91

3.6 (a) 30μm long line scans on flat, uniform 10 nm thick SiO₂ film grown on Si, with fast scan direction parallel to (0° (solid line)) and perpendicular to (90° (dashed line)) the cantilever long axis. (b) Graph of capacitance change versus scan length for 0° (solid line) and 90° (dashed line) scans. .......................... 93

3.7 Illustration of how the bending of the piezoelectric scanner tube during scanning changes the tilt between the cantilever and the sample surface. (a) For a 0° scan, the cantilever is extended toward the probe tip and the cantilever becomes more parallel to the surface. This reduces the average separation between the cantilever and the surface and increases the capacitance. (b) The cantilever is retracted away from the probe tip, causing an increase in the average separation and a decrease in the capacitance. (c) and (d) For a 90° scan, there is a much smaller change in average separation and stray capacitance. .......................... 94

3.8 Schematic of cantilever chip, cantilever, and probe tip during a force calibration curve. (a) Cantilever is undeflected cantilever with the tip out of contact during approach. (b) Cantilever beam deflects downward as attractive/adhesive forces pull the probe-tip onto the surface. (c) Cantilever beam deflects upward as repulsive forces between the tip and sample are caused by further extension of the z-piezoelectric tube. 96
3.9 (a) Force calibration curves and corresponding variations in \( \Delta C \) as the tip approaches and retracts from the surface are shown. The slope of \( \Delta C \) versus \( z \) curve decreases when the tip is in contact. This is due to the separation of the cantilever from the surface being fixed at the probe-tip end (see text). Also note the sudden increase (decrease) in \( \Delta C \) that occurs when the tip jumps to (snaps out of) contact. (b) An enlarged view of the increase in \( \Delta C \) that occurs when the tip jumps to contact is shown.

3.10 (a) Schematic view of a cantilever and tip during a scan across a step. The average separation, \( h \), between the sample and the entire probe assembly decreases abruptly when the tip is scanned from the upper side (a) to the lower side (b) of a step.

3.11 Line traces of (a) surface height and (b) \( \Delta C \) during a 90° scan across a 104 nm deep trench etched through SiO\(_2\) on Si and covered with 10 nm Si\(_3\)N\(_4\). The height scan was flattened to remove an average slope. The \( \Delta C \) scan was flattened to remove the lateral position dependent slope in \( C_{\text{stray}} \) (see text). Dashed line (c) was calculated by multiplying surface height line (a) by -3.6 aF/nm. the slope of the measured \( \Delta C \) versus \( z \) data.

3.12 Force calibration data is used to measure the nm-scale tip-sample capacitance. (a) is a plot of the cantilever deflection measured during approach. Note the sudden snap into contact (i) and the 0 deflection point while the tip is in contact with the surface (vertical dashed line). (b) is a graph of the simultaneously measured capacitance change which occurs during extension. The extrapolated long-range capacitance change caused by the cantilever and probe assembly (ii) is subtracted from the capacitance data to obtain (c), the graph of residual capacitance during extension. Note the sudden increase of capacitance that occurs at snap-on. \( C_{ts} \) is the tip-sample capacitance which is measured when the cantilever deflection is 0.

3.13 (a) A schematic of the system for a spherical capacitor model. Boundary conditions at: (i-iv) are continuity of \( \vec{D} \); tip surface, \( V=1 \). and bottom of film, \( V=0 \). (b) The calculated isopotential lines for a 100 nm sphere on a 50 nm film are shown.
3.14 Capacitance versus dielectric film thickness is plotted for 35, 50, 100, 200, and 300 nm radius spheres. The green line is the calculated capacitance of a parallel plate capacitor of 53 nm radius. .......................... 109

3.15 (a) A schematic cross section of the SiO$_2$/Si trench sample is shown. (b) An AFM topography image of the trench is shown. The full gray scale range is 10 nm. (c) The corrected (see text) capacitance image of the trench is shown. The full scale is $\sim$ 200 aF. .......................... 110

3.16 (a) is the average cross-section of all topography lines in figure 3.15(a). line (i) in (b) is the average of all raw capacitance data lines used to produce figure 3.15(b). The dashed line (ii) is the calculated topography induced parasitic capacitance change while the dotted line (iii) is the corrected capacitance displayed in figure 3.15(b). .................. 112

3.17 The cantilever deflection as a function of piezo-electric scanner tube displacement in $z$ is shown here for extension (blue line) and retraction (red line). Note the different snap-on and snap-off deflections and hence, forces. .................. 115

3.18 The circles represent individual measurements of tip-sample capacitance $C_{ts}$ plotted against the snap-off cantilever deflection and the dashed line is the linear regression (constrained to a 0 intercept) for that data. The correlation is indicative of a relationship between meniscus forces and meniscus effective electrical area. .................. 115

3.19 This is a summary of the affect of environment on tip-sample capacitance. The squares correspond to capacitance measured from the snap-off data and the circles to capacitance measured from the snap-on data 117

3.20 The calculated SiO$_2$ film thickness (red line) is plotted with a schematic film thickness profile (blue line). The dimensions for the schematic are based on ellipsometry measurements of the native (2.5 nm) and thermal (10 nm) SiO$_2$ films. .......................... 119

3.21 The variation in capacitance for 100, 200, and 300 nm radius spheres as a function of height above a 10 nm SiO$_2$ film is presented. Note both the long and short range components of the capacitance. .......................... 121
3.22 A schematic of small and large menisci shows the large, nearly isopotential regions around a sphere that are created by a meniscus. The regions where the electric fields are modified by the presence of a meniscus are shown in yellow; the meniscus is shown in light blue.

3.23 The electric field energy density for a 100 nm sphere on a 10 nm SiO₂ film is shown.

4.1 This is a picture of the 1/2 MOSFET structure that our collaborators at IMEC have made. Notice the large, planar gate region that allows easy access for our grounding electrode and our STM tip. This is not a working device, but is provided to us for the purpose of testing our modification processes.

C.1 The phase shift circuit is two stages of a phase splitter followed by a resistor and capacitor voltage divider. The output of each stage is buffered by a voltage follower. The phase shift introduced in the first stage is fixed while the phase shift introduced in the second stage is variable. The output buffer of the second stage is a fixed gain (2) non-inverting amplifier. The amplitude of the phase shifted voltage is controlled by using a potentiometer as a voltage divider. This is buffered by a unity gain voltage follower.
CHAPTER 1

OVERVIEW

The world we live in is shaped by interfaces and thin films. In biology, the interfaces between lipids and aqueous solutions are responsible for the structure and functions of cellular membranes. In chemistry, interfaces play an important role in catalysis. In electronics, interfaces between different materials and thin films of various materials enable transistors to function. Therefore, understanding the properties of interfaces and films is useful for technological and commercial applications.

Interfaces and thin films are also responsible for a wide variety of interesting physical phenomena. For example, exposed atoms on a cleaved surface rearrange themselves to minimize the surface energy, and form structures that are distinct from the bulk structure [1]. Surface layers can form alloys that are not seen in bulk material [2]. Thin films can be used to fabricate single electron transistors—which exploit the Coulomb blockade seen by electrons tunneling onto a quantum island—for use in developing quantum based capacitance and current standards [3].

This work is directed towards a study of thin, dielectric film electrical properties. In particular, two technologically important systems have been studied: interface defects at the Si/SiO₂ interface; and lubricant films, primarily composed of fully bonded
perfluoropolyethers, that are used to lubricate hard disk drive platters. These lubricant films also show promise for use in micro-electromechanical systems (MEMS).

The first system studied is the charge trapping defect, or interface trap, found at the interface between Si and thermally grown nm-thick silicon dioxide (SiO₂) films grown on the Si. The goal of this work is to make both ballistic electron emission microscopy (BEEM) and charge pumping measurements on the same device. This combination of techniques will allow us to: make nm-scale measurements of interface state formation and hot-carrier transport within working, metal oxide semiconductor field effect transistors (MOSFET) devices: study how interface state formation and device degradation depend on hot-carrier energy and location within the channel; and study how hot carriers move toward source, drain, or substrate as a function of location. These types of nm-scale spatially resolved measurements cannot be made on commercial MOSFETs because BEEM measurements require thin (~5nm), very clean, metal film gates. In contrast, commercial devices are produced with thick poly-crystalline Silicon (p-Si) gates. As a first step in producing a BEEM compatible MOSFET, we have shown that BEEM measurements can be made on metal-oxide-semiconductor (MOS) capacitors that have been subjected to standard semiconductor fabrication processes. Further, through extensive fabrication process development at The Ohio State University (OSU) and The Stanford Nanofabrication Facility (SNF), we have identified the critical technologies required to produce working MOSFETs that are compatible with both BEEM and charge pumping measurements. These efforts have not yet, however, resulted in functional MOSFETs. Therefore, we are engaged in a collaboration with IMEC in Leuven, Belgium. They will produce, and
test working MOSFETs which we will modify and then use in BEEM and charge pumping measurements.

This system is technologically important because these thin SiO$_2$ films are crucial elements in current semiconductor technologies. However, future semiconductor devices will require new dielectric materials because the physical limits of SiO$_2$ films are now being approached. In particular, candidate materials for new dielectrics have shown high densities of electrically active interface defects [4]. These charge trapping interface defect in SiO$_2$ films are responsible for performance degradation of MOS-FETs with time, high voltage, and radiation exposure [5, 2]. An understanding of the physical nature of these defects can help guide the design of new materials to overcome the limits of thin SiO$_2$ films.

The second system under study is the use of capacitance measurements to resolve sub-nm variations in the thickness of thin dielectric films with nm-scale lateral resolution. Towards this goal, we have: developed direct, low-frequency scanning capacitance microscopy (SCM) instrumentation capable of measuring $10^{-18}$F (aF = attofarad) changes in the capacitance between an atomic force microscope (AFM) tip and a sample with a noise level of $0.4 \text{ aF}/\sqrt{\text{Hz}}$: for the first time, quantified and developed means of accounting for changes in parasitic capacitance that occur while scanning an AFM tip; quantified the effective area of the meniscus that forms between the AFM tip and the sample while scanning in air; and measured nm-scale dielectric film thickness variations with $\sim 1\text{nm}$ vertical resolution and $\sim 200\text{ nm}$ lateral resolution.

Variations in nm-thick dielectric films are especially important in nanotribology. For example, the MEMS motor pictured in Figure 1.1 is susceptible to high stiction.
Nanometer scale defects in lubricant films cause high stiction, friction, and early failure of these devices. The lateral uniformity of these monolayer thick lubricant films is critical for consistent tribological performance. Yet, current film thickness measurement technologies cannot be used to assess nm-scale lateral thickness variations because they measure average film thickness over large ($10\mu m$-1 mm diameter) areas.

The reliability of the next generation of hard disk drives is also dependant upon an improved understanding of lateral variations in molecularly thick lubricants. For example, fully-bonded perfluoropolyether lubricant layers are now used in hard disk drive technology to reduce friction and wear at the interface between the slider—which contains the read/write head—and disk surface.
Like the work which underlies it, this dissertation is divided into two principal chapters: Chapter 2 is a review of the Si/SiO$_2$ interface defect studies and Chapter 3 is a description of the SCM work. Chapter 2 is divided into three major sections. The first section contains: a discussion of the physical nature of the Si/SiO$_2$ interface defects, the electrical properties of metal-oxide-semiconductors and MOSFETs, the physics underlying BEEM, and the physics of charge pumping. Then, section 2.3 contains a summary of the BEEM compatible MOSFET fabrication process development work. Likewise, Chapter 3 is also divided into two major sections: section 3.2 is a review of direct, low-frequency SCM instrumentation and parasitic capacitance changes during scanning; and section 3.3 is a presentation of the tip-meniscus-sample capacitance quantification and film thickness variation measurement work. Finally, Chapter 4 provides a general roadmap to future work.
CHAPTER 2

SILICON DIOXIDE INTERFACE STATES

This chapter begins with an overview of the physics and a review of the literature which forms the basis for the combined BEEM and charge pumping measurements which are the goal of this work. It ends with an overview of the fabrication processes used in an effort to produce working, BEEM compatible MOSFETs.

The physical and chemical properties of thin SiO\textsubscript{2} films and the interface between Si and SiO\textsubscript{2} have been studied extensively for the last fifty years because of the technological importance of the Si/SiO\textsubscript{2} interface. Most importantly, thin films of SiO\textsubscript{2} form a key element in metal oxide semiconductor field effect transistors (MOSFETs); the gate oxide (figure 2.1). This oxide forms an insulating barrier between the gate and substrate channel of a MOSFET. It is the very high quality of SiO\textsubscript{2} films and their interfaces with Si that have allowed the development of modern MOSFETs and the computers and consumer electronics that are built from MOSFETs.

The quest for faster and faster electronics has driven the production of smaller and smaller MOSFETs; as they are made smaller switching speeds can be made faster. However, faster switching speeds require higher source-drain saturation currents which in turn require larger gate capacitance [4]. Until now, increased gate capacitance has been achieved by decreasing the SiO\textsubscript{2} film thickness. Now comes the
Figure 2.1: Schematic view of a MOSFET. Two reverse-biased diodes, the source (i) and drain (ii), are separated by a channel (iii). The conductance of this channel is modulated by the potential applied to the gate (iv). The gate oxide (v) is an insulating barrier that prevents current from flowing from the gate to the substrate.

catch. Very thin gate oxides are poor insulators because electrons tunnel through the oxide. There is another way to increase the gate capacitance: increase the dielectric constant of the gate insulating layer by using a different material. One problem with this approach is that the insulating materials tried to date—known as high-k dielectrics—have large defect densities when compared to SiO$_2$ films.

An additional interest in SiO$_2$ films and their interfaces is that the longevity of MOSFETs is adversely affected by the creation of new interface states. Exposure to radiation and operating at high currents can create new oxide interface states.
The mechanisms of defect formation and the location of defect formation in working MOSFETs is not well understood [9, 10, 11, 12].

For these reasons, the desire to understand why the Si/SiO₂ interface has so few defects and how these defects are generated, has driven extensive research into the cause and nature of interface states over the last fifty years. The goal of the work performed for this dissertation is the simultaneous application of two well established techniques. Ballistic Electron Emission Microscopy (BEEM) and Charge Pumping, to study the energetics and position dependence of interface state creation in working MOSFETs. BEEM experiments reveal details about charge and barrier potentials at buried interfaces with nm-scale spatial resolution [13, 14, 15]. In a BEEM experiment, energetic electrons are injected from a scanning tunneling microscope (STM) tip through the gate metal, gate oxide, and into the substrate where they are detected as a current. The electrons only transit the oxide if the energy of the injected electrons is greater than the barrier height of the oxide. Charge pumping experiments reveal details about interface trap density of states, charge capture and emission cross-sections, and capture and emission time constants in an entire device. In a charge pumping experiment, the MOSFET source and drain are reverse biased with respect to the substrate and voltage pulses are applied to the gate metal. Gate voltage pulses move interface traps in and out of equilibrium so that they alternately trap minority carriers supplied by the source and drain and emit minority carriers into the substrate. The minority carriers are removed from the substrate by recombination with majority carriers. As a result, a DC current flows from the source and drain to the substrate. It is this DC current that is measured in charge pumping experiments (section 2.2.4).
A combined BEEM and charge pumping experiment requires specialized MOS-FETs that have thin metal gates (~ 5 nm) and thin oxide films (< 10 nm). No such devices are available commercially so we have attempted to develop a BEEM compatible fabrication process. While we have made substantial progress towards producing these devices, we are still encountering problems with gate oxides that are not sufficiently insulating. Hence, the Pelz group is now engaged in a collaborative effort with IMEC. They will provide mostly complete—and tested—devices. The Pelz group will remove the existing gate, deposit a thin metal gate, and perform BEEM and charge pumping experiments.

In sections 2.1-2.2.4, I will review the literature on the physical nature of the interface defects, the electrical characteristics of the Si/SiO₂ interface, the physics of BEEM, and the physics of charge pumping. Then, in section 2.3, I will describe and critique the fabrication procedures used in our attempts to produce a working, BEEM compatible MOSFET.

2.1 Physical properties of thin SiO₂ films on Si

A physical picture of the Si/SiO₂ interface is helpful in understanding the origin of interface traps. In this section, the geometry and crystallinity of the interface is discussed in terms of TEM data, x-ray diffraction data, and Monte Carlo simulations of a random, continuous network of bonds. Next, the chemical state of Si at the interface is discussed in terms of XPS data, cathodoluminescence data, and Molecular Dynamics simulations. Finally, the relationship between defects at the Si/SiO₂ interface and interface states is discussed via overviews of EPR data and interface state passivation and generation studies.
The interface is the transition region between the crystalline Si and the amorphous SiO$_2$. Determining the nature of the interface is so difficult because it is buried. Further, the disordered and insulating nature of the SiO$_2$ layer makes interpretation of traditional structural techniques such as transmission electron microscopy (TEM) difficult and ambiguous. No single experimental technique yields unambiguous results, and hence, a wide variety of experimental techniques has been applied to the problem. X-ray photoemission spectroscopy (XPS) and electron paramagnetic resonance (EPR) have been used to examine the bond chemistry at the interface. Transmission electron microscopy (TEM), low energy electron diffraction, and x-ray diffraction have been used to study the atomic geometry of the interface. Capacitance measurements, charge pumping measurements, and ballistic electron emission microscopy (BEEM) have been used to study the electrical properties of interface traps and the oxide. Hot-Carrier injection, internal photoemission, irradiation, and annealing experiments have been used to study the formation and passivation of interface defects. In addition, a variety of theoretical methods—including molecular dynamics (MD), monte carlo simulations (MC), and density functional theory—have been used to try and elucidate the atomic geometry and bonding states of the Si/SiO$_2$ interface.

The basic picture of the Si/SiO$_2$ transition region that emerges from this data is that it is only a few monolayers across [16, 17, 18. 10] and that the a-SiO$_2$ region is relaxed and not strained [17, 18]. Further, nearly all of the Si atoms at the interface are bonded to four other atoms. It is this statement that explains the very low density of interface states ($10^{10}$–$10^{12}$ states/cm$^2$) measured with electrical techniques [19, 20].
2.1.1 Geometry of Interface Structures

The density of atoms in Si is about 2.2 times the density of atoms in SiO\(_2\). In order for the Si/SiO\(_2\) interface to be relaxed and the interfacial Si bonds to be saturated: the structure of the Si must be perturbed, the structure of the SiO\(_2\) must be perturbed, or both must be perturbed.

Transmission electron microscopy is used in lattice imaging mode to obtain images of the Si/SiO\(_2\) interface. In this mode, a beam of \(~100\text{keV}\) electrons is projected through a thinned sample. The diameter of the beam is perpendicular to the sample surface; the length of the beam is parallel to the interface and is directed along a specific crystallographic direction. This creates shadows of the lattice in the TEM image: these may be used as an internal magnification standard. TEM images, then, are projections of the three dimensional sample onto a plane. The resulting image reflects the average properties of the atoms in a given column of sample.

TEM studies of the Si/SiO\(_2\) interface show a flat, well defined interface and an abrupt transition from crystalline Si to amorphous SiO\(_2\). For example, Carim and Sinclair studied several thicknesses of thermal oxides grown on (100)Si [21]. They reported surface roughness that, after an initial increase, decreased with increasing oxide thickness and attribute this to initial non-uniformities in surface oxidation. The maximum roughness reported was 14 Å; more typical values were 6–8 Å. They attributed the roughness to differential oxidation at step edges that is relaxed—in the thicker oxides—via viscous flow during oxidation. Note that these results may be interpreted as an atomically abrupt transition from crystalline Si to amorphous oxide on a microscopically rough surface. Another TEM study was done by Hollinger
et. al. [22] on device quality thermal oxides. They too report abrupt interfaces on a microscopically rough surface.

Grazing incidence x-ray diffraction studies complement the TEM data. They are done by directing a beam of x-rays at angles near that required for total external reflection at the surface of a sample. The oxide does not contribute to the diffraction pattern because the random atomic positions cause random phase shifts that result in destructive interference. Because the angle of incidence is low, only the surface layers of the crystal are imaged. Periodic, two-dimensional surface lattices cause lines to appear in the diffraction image. The spacing and intensity of these lines are then related to the surface pattern. As with TEM, x-ray diffraction experiments give an average of the properties seen across the surface. Aperiodic structure variations cause destructive interference and cannot be observed.

Grazing incidence x-ray scattering experiments have also been used to look for order in the transition layer. Renaud et. al. [23] measured x-ray diffraction patterns from the surface of normal (~ 0.25°) and very low miscut (~0.5°) (001)Si wafers with a native oxide layer (~ 16 Å). They grew the native oxide layer in dry oxygen at one atmosphere pressure. They found an interfacial phase that is coherent with the Si substrate and less than 5.5 Å thick. A diffraction pattern indicative of superposed 2x1 and 1x2 symmetry was found on the low miscut samples. This could be caused by cells aligned on the alternating direction steps found on (001)Si. Further evidence of this is the finding that the coherence length is proportional to the separation between steps and the absence of the pattern from the normal miscut sample.

One possible explanation for this surface symmetry comes from Monte Carlo simulations of a continuous random network model of the Si/SiO₂ interface performed
by Tu and Tersoff [24, 25]. Their algorithm starts with a perfect interface between crystalline Si and α-cryostobalite, a highly strained configuration. The energy of this configuration is calculated from a valence force model that depends only on bond topology. Next, a Monte Carlo step is performed by randomly switching bonds. The energy of this new configuration is calculated and the move is accepted with a probability given by

\[
P = \min[1, e^{-\delta E / k_B T}],
\]

where \( \delta E \) is the change in configuration energy, \( k_B \) is the Boltzmann constant, and \( T \) is the temperature.

The key result of this study is that an oxygen bridge bond appears between Si atoms on the interface. This allows the bonds at the surface to be saturated while allowing the oxide density at the surface to be the same as the bulk oxide density. This is seen as a reduction in energy as the number of bridge bonds increases. A consequence of the bridge bonds is that a stripe phase appears at the interface. This stripe phase is consistent with that seen by Renaud et al. [23].

### 2.1.2 Bonding and Stoichiometry at the Interface

While TEM and x-ray diffraction provide a picture of the position of the atoms on the surface, they do not provide information on the bonding between the atoms. For this, surface sensitive chemical techniques are required. XPS and EPR are two such techniques. In XPS, x-rays incident on the sample knock core level electrons out of their orbitals. If the emitted electron is close enough to the surface, it escapes from the sample and its energy may be measured. The difference between the electron
energy and the gamma-ray energy provides information about the local chemical environment. Peaks in the spectra of Si/SiO₂ systems then reflect the oxidation states present in the sample. For example, in SiO₂, the Si oxidation state is Si⁺⁺⁺: each Si is bound to four O atoms. In contrast, in Si, the Si oxidation state is Si⁰: each Si is bound to four other Si atoms. Electrons which are emitted from Si atoms with different oxidation states see very different potentials as they escape: they lose different amounts of energy as they escape. One important limit of XPS is sensitivity. It cannot detect states with a surface density of less than 10¹² – 10¹³ cm⁻²[18]. This is higher than the interface state density determined by electrical measurements. Hence, while it can provide information about the oxidation state of Si atoms at the interface, it cannot provide information about the interface states.

XPS measurements from Si/SiO₂ samples have been used to measure Si oxidation states in the oxide and at the interface. Grunthaner et. al. measured oxidation states of thermal oxides grown on (100)Si with XPS and depth profiling of the oxide by successive wet-chemical-etches to remove oxide [26]. They used the XPS spectra due to Si 2p and O 1s lines to calculate oxidation states. They used a Fourier transform to deconvolve portions of the spectra that are common to both lines. The remaining lines indicate the presence of all intermediate oxidation states of Si near the interface and that the region within about 30 Å of the interface contains substoichiometric oxide. Further analysis was used to support a model of the interface as a continuous random network of four to eight member rings of SiO₄ tetrahedra joined by oxygen bridges.

Himpsel et. al. made high resolution XPS measurements on (111) and (100) Si wafers with thin SiO₂ films grown at atmospheric pressure in oxygen [27]. These
conditions yield native oxide films. The spectra were processed to deconvolve the peaks due to the intermediate oxidation states: $\text{Si}^{1+}$, $\text{Si}^{2+}$, and $\text{Si}^{3+}$. First, the secondary electron background curve, measured at a 10-eV lower energy shift, is subtracted from the Si 2p core-level shift spectrum. After this, a small background due to energy losses from core lines is subtracted. The peaks remaining are interpreted as being due to the intermediate oxidation states. This process was repeated for different incident energy x-rays. Quantitative depth information was not obtained because of a resonance in $\text{Si}^{3+}$ absorption cross section at 130 eV incident energy. However, the intermediate oxidation states were only observed at the higher energies. This indicates that they are located at or near the surface and not in the bulk of the $\text{SiO}_2$.

The intensity ratios of the various peaks, when corrected for photoemission cross-section and electron escape depth, can be used to estimate the abundance ratio of the oxide states. $\text{Si}^{1+}$, $\text{Si}^{2+}$, and $\text{Si}^{3+}$ are all found to be at about the same abundance [27]. It should be noted that these calculations are sensitive to electron escape depth estimates. The cross-sections can be measured in bulk Si and bulk $\text{SiO}_2$, so, they are fairly reliable.

Molecular dynamics (MD) simulations were performed by Pasquarello, Hybertsen, and Car to study model interface structures used to interpret XPS data [28, 29, 30, 31]. They calculated Si 2p core-level shifts by attaching tridymite [28, 30] and crystabalite [119, 127] to (100)Si. Pseudopotentials derived from local density approximations in density functional theory (LDA-DF) [32, 33] were used in force calculations and classical molecular dynamics was used to relax the system. The systems were started with either tridymite or $\beta$-cristobalite to (100)Si. The forces between atoms were
calculated via LDA-DF. and the atom postions were moved under this force with a
Verlet integration [32. 33]. This process was repeated until the structures relaxed.
They found that core level shifts are primarily sensitive to nearest-neighbor oxygen
atoms and that shifts due to second nearest-neighbor oxygen atoms are negligible.
This means that the XPS measurements are not sensitive to bond topology. Further,
all three intermediate oxidation states of Si are found at the interface while interface
states in the Si band gap were not found.

2.1.3 Defects and the Nature of Interface States

From the comparison of structural studies and electrical measurements of interface
states, it is clear that interface states are associated with defects rather than the
"perfect" interface. EPR is a particularly good way to investigate the structure
of defects. Its sensitivity can be, with extreme care, sufficient to detect as few as
$10^{10}$ defects/cm$^2$. This is of the same order of magnitude as the defect density in the
best device quality oxides. In EPR studies, the sample is placed in a slowly varying,
high magnetic field. A microwave frequency magnetic field is applied perpendicular to
this field. An isolated electron in this environment would have a resonance frequency
given by

$$h\nu = g_0\beta H.$$  \hspace{1cm} (2.2)

where $h$ is Planck's constant. $\nu$ is the microwave frequency. $g_0$ is the electronic g-factor
(2.002319), $\beta$ is the Bohr magneton. and $H$ is the magnetic field strength. In atoms
and molecules, spin-orbit coupling and the presence of paramagnetic nuclei alters the
resonance frequency. The change in resonance frequency is known as the chemical
shift; it is a reflection of the local chemical environment of the unpaired electron. In solids, spin orbit coupling is suppressed for the ground state. However, excited states do contribute to spin orbit coupling. Perturbation theory is used to calculate the coupling: the \( g_0 \) factor is replaced in 2.2 with

\[
g_{ij} = g_0 \delta_{ij} - 2\lambda \sum_k \frac{\langle \alpha | L_i | k \rangle \langle k | L_j | \alpha \rangle}{(E_k - E_\alpha)}.
\]

(2.3)

where |\( \alpha \rangle \) is the ground state wavefunction with energy \( E_\alpha \). \( \lambda \) is the atomic spin-orbit coupling constant. \( L_i \) and \( L_j \) are angular momentum operators for the relevant \( x \), \( y \), and \( z \) directions. and the sum is over all excited states [16].

EPR studies of the Si/SiO\(_2\) system have identified several defects related to the interface and interface states. They include several P\(_b\) centers, several E\(^-\) centers, a peroxo radical (O\(_2\)Si\(^-\)), and non-bridging-oxygen hole centers [17, 16]. The P\(_b\) centers and E\(^-\) centers are the two classes of defect that have been most identified with the formation and passivation of electronic interface states. The P\(_b\) centers have been identified as Si dangling bonds at the Si/SiO\(_2\) interface. The E\(^-\) centers have been, tentatively, identified as oxygen vacancies in the oxide. They are involved in interface state generation through the trapping of hot-holes injected into the oxide [16, 17].

Three types of P\(_b\) defects are frequently observed. In (111)Si/SiO\(_2\) systems, a single type of P\(_b\) defect is observed. In contrast, in (100)Si/SiO\(_2\) systems, two types are commonly observed: P\(_{b0}\) and P\(_{b1}\) [16]. All have been identified as Si dangling bonds in the following manner. \(^{29}\)Si is a spin 1/2 nucleus with a natural abundance of 5%. The \(^{28}\)Si nucleus is spin zero. When both nuclei are present in their natural abundances, they produce a unique three line pattern. The central peak, with 95% of the intensity, is due to \(^{28}\)Si. The two side lobes, each with 2.5% of the intensity, are due to the two resonances of the \(^{29}\)Si. This pattern is observed for all of the
Pb defects. Further, EPR measurements on $^{17}$O enriched oxides do not show line broadening due to nearest neighbor interactions of the nuclear moment of $^{17}$O with the Pb defects. This indicates that the Si is not bonded to O.

The Pb defect found in the (111)Si/SiO$_2$ system is the best characterized of the defects. Reviews by Lenahan and Conley [16] and Helms and Poindexter [17] discuss the history of this characterization. Their conclusions are as follows. The (111)Si/SiO$_2$ Pb defect has an axially symmetric $g$-tensor with its symmetry axis in the (111) direction. This is, of course, the direction dangling bonds would point in the unreconstructed (111)Si system. Further, comparison with the much stronger signals from similar defects in bulk Si, shows that the Pb defect is a trivalent Si. When the oxide is etched off with HF, the signal disappears. Therefore, it must be within a few monolayers of the Si surface or in the oxide. The Pb defect is sensitive to carrier density at the interface; therefore, it must be within a few monolayers of the interface. Theoretical calculations of the hyperfine splitting for models of the (111)Si/SiO$_2$ system are in qualitative agreement with this model [11]. When taken together, these data provide strong evidence that the Pb defect in (111)Si is a dangling Si bond at the interface.

The story is more complicated for (100)Si/SiO$_2$ systems. In these systems, identification of the Pb$_0$ and Pb$_1$ defects is less sure. Helms and Poindexter [17] discuss the resolution of the overlapping spectra from these defects into two components. The $g$-tensor they derive for the Pb$_0$ defect is almost the same as that for the $g$-tensor for the Pb defect on the (111)Si/SiO$_2$ interface. The symmetry axis in this case is again in the (111) direction. This is again consistent with the dangling bonds on the
unreconstructed (001)Si surface. The $P_{b1}$ defect $g$-tensor does not have the same
degree of symmetry, and its detailed structure is uncertain.

The structural and chemical data do point to one solid conclusion: interface states
are not the consequence of the overall structure of the interface. Calculations on defect
free interfaces do not result in interface states, and the density of interface states in
device grade oxides is in agreement with the defect density rather than the unit cell
density. This leads one to examine defects as the cause of interface states. EPR
studies of the Si/SiO$_2$ system points to the $P_{b1}$ class of defects which are ascribed to
dangling Si bonds at the Si/SiO$_2$ interface as being the structural basis for interface
states.

### 2.2 Electrical Characteristics of SiO$_2$ on Si

It is the electrical characteristics of Si/SiO$_2$ interface defects—or interface traps—
that are of interest in this study. We concentrate on three methods of measuring
interface trap electrical properties: capacitance, BEEM measurements, and charge-
pumping. Capacitance measurements reflect the changes trapped charges induce
in semiconductor capacitance-voltage characteristics and the different interface trap
responses to high and low signal frequencies. BEEM measurements sense alterations
in local barrier height caused by the charge trapped by interface states. BEEM may
also be used to inject charge at known energies with high spatial resolution to create
or alter interface states. Charge pumping measurements yield: the total number of
interface states; the density of states as a function of energy; interface state emission
and capture cross-sections; and interface state emission and capture time constants.
2.2.1 Metal–oxide–semiconductor capacitor physics

A digression into the basic electrical properties of the thin SiO\textsubscript{2} film on Si system will help explain how each of these techniques probes interface traps and what can be learned from each technique. A metal–oxide–semiconductor capacitor is the most basic structure used for evaluating electrical properties of interface traps. A metal film is deposited on a thin oxide film that was grown on Si. A schematic of the separate components of a MOSCAP are shown in figure 2.2(a). Here both vertical and horizontal axes are space axes. A schematic of the band diagrams is shown in figure 2.2(b); here the vertical axis is electron energy and the horizontal axis is a space axis perpendicular to the Si/SiO\textsubscript{2} interface with its origin at the metal–SiO\textsubscript{2} interface.

The fermi levels of the gate metal and the Si—p-type in this diagram—are referenced via the the energies required to remove an electron from the metal ($eW_g$) and Si surfaces ($eW_{Si}$) to the vacuum level. The "band structure" of the oxide is depicted as a rectangle. Now consider figure 2.2(c) where the gate is electrically connected to the substrate. Before being connected, the Fermi levels, with respect to the vacuum level, in the metal and the Si are different: after being connected, the Fermi levels in the metal and the Si are the same and the system is in equilibrium. The equality of the Fermi levels is achieved by transferring charge from the gate to the substrate until the system is in equilibrium. This charge transfer causes a voltage across the oxide because no current can flow through the oxide. Further, unlike a metal, the free charge carrier density in the Si depends on the local potential. This allows an electric field to exist in the semiconductor and the voltage drop occurs across the series combination of the oxide and a small region in the semiconductor near the Si/SiO\textsubscript{2}
interface. This voltage drop in the Si is the origin of the conduction and valence band bending pictured in figure 2.2(d). In this case, a depletion layer is formed where free charges—holes for p-type Si—are swept away from the interface. The negatively charged dopant ions remain fixed in the lattice so the depletion region is negatively charged. This charge distribution is shown at the bottom of figure 2.2(d). There is one other consideration: what is the relationship between the oxide band edges and the gate metal and semiconductor band edges? These modified work functions, $\phi_g$ for the gate metal and $\phi_S$, for the Si, are the minimum energies between the respective Fermi energies and the oxide conduction band minimum [4]. The modified work functions are determined by the Si/SiO$_2$ interface properties and are measured experimentally, often with photoemission spectroscopy [5].

The charge redistribution discussed above results from the difference between the gate metal and Si modified work functions. Now consider the effect of applying a voltage between the gate and the substrate. The metal and Si fermi levels are no longer aligned; they are split by $eV_a$ where $e$ is the electronic charge and $V_a$ is the applied voltage. A MOSCAP with a DC bias is in equilibrium because the oxide prevents current flow. Charge is, however, transferred through the external connection to shift the relative Fermi levels and re-establish equilibrium.

There are three different regimes with a DC bias voltage applied: accumulation, depletion, and inversion. These regimes, for p-type Si, are depicted in figure 2.3(a), (c), and (d). Accumulation occurs when a negative voltage applied to the gate attracts holes to a thin layer at the Si/SiO$_2$ interface. The flatband condition occurs when the gate voltage just offsets the difference between the metal and Si work functions (figure 2.3(b)). There is no accumulation of charge at the interface and hence, no
Figure 2.2: Schematic view of (a) MOS capacitor components, (b) the band diagram for the separate components, (c) the assembled MOS capacitor, and (d) the band diagram for the MOS capacitor with the gate connected to the substrate.

Voltage drop across the oxide. Depletion occurs when a positive voltage applied to the gate drives holes away from the interface and leaves a region devoid of free charge; the fixed, negatively charged dopant atoms remain in place leaving a charged region. Increasing the applied voltage increases the width of the depletion region until the onset of inversion. Inversion occurs when the applied voltage brings the conduction band minimum at the Si/SiO₂ interface below the extrinsic Fermi level.
and the conduction band at the surface fills with electrons. Further increases in applied voltage bring more electrons to the surface but does not increase the thickness of the depletion region. The onset of inversion occurs when the surface potential is approximately twice the difference between the intrinsic and extrinsic fermi level [5]. For n-type Si, the above description holds with electrons and holes, and positive and negative, exchanging roles.

Figure 2.3: Band bending diagrams for (a) accumulation, (b) flatband, (c) depletion, and (d) inversion for a p-type Si MOS capacitor.
Interface traps change this picture of the MOSCAP response to voltage applied to the gate by adding a layer of charge to the Si/SiO$_2$ interface. This charge acts to shield the Si from potential changes caused by gate voltage changes. Fixed charge densities would simply shift the gate voltages required to place the MOSCAP in accumulation, depletion, or inversion. However, the density of interface trapped charge is a function of the Si surface potential so the degree to which the surface is shielded from gate voltage changes as a function of the Si surface potential. This results in a broadening of the gate voltage range required to swing the Si surface from accumulation to inversion.

The above band-bending description of the MOSCAP with interface traps is useful for explaining the capacitance-voltage (CV) characteristics of MOS SYSTEMS. The voltage dependence of charge distribution in the surface means that MOS structures exhibit voltage dependent changes in capacitance. Hence, for these systems, differential capacitance, defined as $C' \equiv \delta Q/\delta V$, is measured. Low frequency CV characteristics will be considered first. Low frequency means that applied signals are slow enough that the minority carriers are always in equilibrium. In this case, the low-frequency equivalent circuit for this system is two capacitors in series. The metal-oxide-Si surface forms one capacitance $C_{ox}$ while the Si surface-depletion region-bulk Si forms the other capacitance $C_{Si}$. There are two voltages that are relevant when describing MOSCAP-CV characteristics: the DC gate bias voltage and the small, AC gate voltage used to measure the system capacitance. Figure 2.4 depicts the equivalent circuit, the applied voltages, and a low frequency, n-type Si MOSCAP CV characteristic. There are four regions in the low frequency MOSCAP CV characteristics depicted in figure 2.4(b): (i) accumulation, where the surface is pinned by the
Figure 2.4: (a) A schematic of a MOSCAP is a depiction of the applied voltages and equivalent capacitances for low frequency capacitance measurements. (b) A schematic low frequency CV characteristic for a MOSCAP showing the (i) accumulation, (ii) depletion, (iii) inversion onset, and (iv) deep inversion regimes. The blue line is a MOSCAP without interface traps: the red line depicts the broadening and shallowing of the CV characteristic as described in the text.

accumulation of electrons (majority carriers) at the Si/SiO₂ interface; (ii) depletion, where electrons are driven from the surface and a depletion layer forms; (iii) inversion onset, where the conduction band at the surface crosses the Fermi level and holes (minority carriers) supplied by carrier pair generation start to form an inversion layer; and (iv) deep inversion, where the Si surface potential does not change in response to gate voltage changes. The blue and red lines in figure 2.4 depict the CV characteristic in the absence and presence of interface traps. As described above, interface traps introduce a voltage dependent charge layer at the Si/SiO₂ interface that shields the Si surface from changes in gate voltage. This broadens the low frequency CV characteristic in the presence of interface traps.

High frequency capacitance measurements result in a very different CV characteristic because the majority carriers respond to gate potential changes but the minority
carriers do not. Hence, the minority carriers are not in equilibrium with the majority carriers. Schematics of the high frequency CV equivalent circuit and characteristic are presented in figure 2.5. There are different regimes seen in the high frequency CV characteristic depicted in figure 2.5: (i) in accumulation the majority carriers can follow the signal and the surface potential remains pinned; (ii) in depletion the electrons (majority carriers) follow the signal and are driven from the depletion region; (iii) at inversion onset and (iv) in deep inversion the holes (minority carriers) cannot follow the signal so no inversion layer forms and the depletion layer continues to grow with increasing negative bias. The equivalent circuit for high frequency MOSCAP measurements is depicted in figure 2.5. It consists of an oxide capacitance that is in series with the parallel variable resistance $R_{Si}$ variable capacitance $C_{Si}$ combination due to the Si. Both $R_{Si}$ and $C_{Si}$ are functions of the Si surface potential ($V_{Si}$). To understand the physical basis for the resistance consider the fate of thermally generated electron-hole pairs in the depletion region. After generation the potential drop across the depletion region separates the pairs. Now, before recombination occurs, the high frequency signal changes the width of the depletion region. This causes some of the pairs that were generated in the depletion region to recombine in the substrate. These pairs were generated out of equilibrium with the substrate so they deposit energy into the substrate. This effect produces a dissipative current that is in phase with the applied AC voltage. A similar phenomenon occurs with interface traps. Interface traps that are out of equilibrium capture and emit hot carriers out of phase with the applied voltage so they transfer excess energy to the lattice. Hence, interface traps add another complex impedance to the equivalent circuit [34, 19].
Figure 2.5: (a) A schematic of a MOSCAP is a depiction of the applied voltages and equivalent capacitances for high frequency capacitance measurements. (b) A schematic CV characteristic for a high frequency MOSCAP showing the (i) accumulation, (ii) depletion, and (iii) inversion regimes. The blue line is a MOSCAP without interface traps; the red line depicts the broadening and shallowing of the CV characteristic as described in the text.

### 2.2.2 Ballistic Electron Emission Microscopy

Ballistic electron emission microscopy (BEEM) is an enhancement to scanning tunneling microscopy (STM) that adds the ability to probe buried interfaces. In particular, BEEM probes the energy barrier at the buried interface. The MOSCAP is an ideal structure on which to do BEEM measurements. BEEM works in a MOSCAP as follows. Electrons are emitted by the STM tip, tunnel through the vacuum and into a thin metal film. Most of these electrons scatter in the metal, equilibrate at the metal Fermi level, and are collected as the STM feedback current. However, some of the electrons do not scatter and proceed through the film. If they are higher in energy than the oxide "conduction band minimum" and the oxide film is thin, they travel through the oxide film and are collected in the Si substrate as BEEM current. Figure 2.6 is an energy band diagram of a BEEM experiment on an n-type MOSCAP.
Figure 2.6: (a) A schematic of the band structure during a BEEM measurement. The STM tip bias $V_{tm}$ controls the energy of the tunneling electrons while the sample bias $V_{ms}$ controls the voltage drop across the oxide barrier.

It includes a bias voltage that may be applied between the Si substrate and the metal film to vary the oxide energy barrier. When the Si bias voltage is positive, the voltage across the oxide barrier reduces the probability that an electron will scatter from the oxide to the metal: the BEEM current is increased [35, 15]. In contrast, when the Si bias voltage is negative, the voltage drop across the oxide increases the oxide barrier height. More energy is now required for an electron to pass over the oxide barrier. Hence, by varying the Si bias voltage, BEEM spectroscopy of the local oxide barrier height may be done.

Note that charge trapped in the oxide or in interface traps changes the local oxide barrier height [36, 37, 38, 35]. Further, BEEM may be used to inject charge into the
oxide and interface states. This is one way we intend to use BEEM in conjunction with charge pumping experiments. We will measure the interface state properties and then inject high energy electrons into the oxide and Si. We will then reevaluate the interface trap properties with charge pumping. The differences in the interface trap properties could then be attributed to the creation or modification of an interface trap by hot electron injection at known energy and location.

There is another complexity to take into account when interpreting BEEM spectroscopy measurements: image force lowering [15, 39, 35]. An electron in the oxide film is attracted to the grounded metal plate. This produces a negative potential which decreases as the electron travels through the oxide. Now, when a positive voltage is applied to the substrate, a linearly decreasing potential is formed in the oxide. The sum of these two potentials is less than the metal gate modified work function \( \phi_m \). Consequently, the BEEM threshold voltage decreases.

### 2.2.3 MOSFET Operation

An understanding of the details of MOSFET operation is required before charge pumping can be explained. The following is a description of the low frequency operation of an enhancement mode, n-channel MOSFET. It is composed of two heavily doped n-type regions separated by a lightly doped p-type channel. Over the p-type channel is a thin, high quality insulator, often SiO₂. Figure 2.1 is a drawing of the cross section through an n-channel MOSFET. In operation, a bias voltage is applied to the gate \( V_g \) and a supply voltage \( V_{dd} \) is applied between the drain and source. Typically, the source and substrate are both connected to ground. When \( V_g \) is less
than the flat band voltage, the channel surface is in accumulation and holes are concentrated at the surface. The source-substrate and drain-substrate pn junctions are both reverse biased so no current can flow between the source and drain. For $V_g$ above the threshold voltage $V_T$, the channel surface is in inversion and electrons are concentrated at the surface. Now, at the Si/SiO$_2$ interface, the source-substrate and drain substrate junctions are no longer pn junctions. Electrons flow through the channel from the source to the drain. The conductance of this channel as a function of changes in gate bias is the transconductance. The drain-source voltage $V_{dd}$ also has an impact on device operation. The drain current $I_{dd}$ is given by

$$I_{dd} = \frac{W \mu C_{ox}}{L} \left[ (V_g - V_T) V_{dd} - \frac{V_{dd}^2}{2} \right].$$

(2.4)

where $W$ is the width of the MOSFET, $L$ is the length of the channel, $\mu$ is the average electron mobility, and $C_{ox}$ is the gate-oxide capacitance [6]. For small values of $V_{dd}$, the drain current increases nearly linearly with $V_{dd}$. At higher $V_{dd}$ the channel surface potential near the drain is reduced, the channel depth—into the Si—decreases, and the drain current increases more slowly with $V_{dd}$. Finally, the drain current saturates when $V_{dd}$ is high enough that the channel near the drain is no longer inverted. The drain current is not cut off because the electrons coming from the source are injected into the depletion layer. They are swept through the depletion layer by its electric field.

An analysis of the potentials in the MOSFET relative to the Fermi level is useful during analysis of the affect of interface traps on MOSFET performance. Figure 2.7 is a set of diagrams showing the variations in electronic energy levels at the Si/SiO$_2$. 

30
interface during MOSFET operation and the effect on equilibrium occupancy of interface traps. In figure 2.7a, the drain–source voltage \( V_{dd} \) and the gate voltage \( V_g \) are zero. All of the acceptor traps below the Fermi level, depicted as a dashed line, are full while those above are empty. The MOSFET channel surface is in accumulation so the source–substrate and drain–substrate diodes are both reverse biased. There is no conducting channel. Now consider a positive gate voltage while maintaining \( V_{dd} \) at zero. The band edges bend down and the surface below the gate moves into depletion.
For a suprathreshold gate voltage, as in figure 2.7b, the intrinsic level bends below the Fermi level and the surface is in inversion. The surface below the gate now forms a conducting channel. A voltage applied between the source and drain \( V_{dd} > 0 \) now causes a current flow from the drain to the source. This steady state condition is depicted in figure 2.7c.

Now, consider the affect interface states in the surface conduction channel have on MOSFET operation. In equilibrium, interface states above the Fermi level are empty, while those below the Fermi level are full. When the gate bias is swung from accumulation, first into depletion, and then into inversion, the interface state energies are lowered relative to the Fermi level. As the trap energies are lowered with respect to the Fermi level, more of the states emit holes into the valence band. These holes are swept into the substrate by the electric field. The charge captured and emitted by these traps acts a voltage dependant capacitance in series with the oxide capacitance, and in parallel with the depletion layer capacitances [34].

In contrast to the equilibrium situation, when the gate channel is brought from accumulation to inversion faster than the interface traps can follow, the traps initially remain empty but start to be filled with electrons from the surface inversion layer. Now, as the gate channel is brought back to accumulation, the traps are again out of equilibrium, this time with trapped electrons. The trapped electrons are gradually emitted into the valence band. This is equivalent to trapping a hole. In effect, the interface trap finite emission rate places a low-pass RC circuit between the source/drain regions and the gate channel. The RC time constant is dependant upon the emission rate [19]. The complex admittance of these traps may be calculated with Shockley Read-Hall statistics [40, 41]. Combining the equivalent admittances allows one to
calculate the surface potential as a function of gate voltage. It is the non-equilibrium behavior of interface traps that is exploited in charge pumping measurements.
2.2.4 Charge pumping physics

Charge pumping experiments reveal the following details about interface traps: density of states, charge capture and emission cross-sections, and capture and emission time constants for the population of traps in an entire device. Charge pumping measurements are made by applying a train of pulses to the gate of a MOSFET and measuring the DC current through the substrate. The DC current arises as the pulses alternately swing the channel surface between accumulation and inversion and interface traps alternately fill with minority carriers from the source and drain regions and emit the minority carriers into the substrate. In the substrate, the minority carriers recombine with the majority carriers. The result is the DC charge pumping current $I_{cp}$.

Charge pumping is an effect of non-equilibrium trap dynamics as the MOSFET gate is swung from accumulation to inversion and back. The gate voltage swings the gate channel from accumulation to inversion faster than the traps can empty. This traps majority carriers in the inverted gate channel. They are emitted over time and recombine with minority carriers from the source and drain. The gate voltage then swings the gate channel from inversion to accumulation. Again, the change in state occurs faster than the traps can empty of minority carriers and the minority carriers remain trapped. They are emitted over time into the valence band where they recombine with majority carriers. This results in a DC current through the substrate.

Density of states as a function of energy can be determined because the emission time of a trap is energy dependant. The closer in energy a trap is to the conduction or valence band edge, the shorter the emission time to that band edge is. By selecting the voltage pulse rise/fall time, one selects a subset of interface traps that are kept
out of equilibrium during the voltage transitions. Only the interface traps at these selected energy levels contribute to the charge pumping current. Selecting a series of voltage pulse rise/fall times allows one to build a density of states function $D(z)$ for the interface trap population for energies $z$ from a few $kT$ above the valence band edge to a few $kT$ below the conduction band edge.

Brugler and Jespers [12] performed the first charge pumping experiment. They applied a train of square pulses to the gate of a MOSFET device. The source and drain were maintained at a reverse bias voltage and the DC current through the substrate was measured. Figure 2.8 is a schematic of a charge pumping experiment performed on an n-channel MOSFET. Brugler and Jespers found that the charge pumping current was linearly proportional to frequency, that it decreased with increasing source–drain reverse bias, that it saturated with increasing pulse amplitude, and that the pulse voltage must sweep the surface between accumulation and inversion. Also, they noted that the charge pumping current was reduced when the applied pulse was triangular rather than square.

Brugler and Jespers assigned the source of the charge pumping current by examining the possible sources of charge. They noted that surface generation, generation in the depletion region, and diffusion from the bulk followed by a sweep across the depletion region are all too slow. However, the source and drain regions provide a ready source of minority carriers. In their description, charge pumping occurs as follows. As the surface is pulsed into inversion, minority carriers flow from the source and drain to fill the inversion layer. Some of these carriers recombine with majority carriers from the substrate. As the surface is pulsed back into accumulation, minority carriers are swept back into the source and drain. The trapped minority carriers
then recombine with majority carriers from the substrate. This results in a loss of charge with each cycle. Multiplying this charge by the pulse train frequency and the gate area yields the charge pumping current, $I_{cp}$. This description of charge pumping implies that charge pumping is a surface phenomenon because inversion only occurs at the surface.

Further experiments lead Brugler and Jespers to assign two components to the charge pumping current: a geometric component and a fast interface state component. The geometric component arises, they argued, when majority carriers returning from the substrate reach the surface before all of the minority carriers are able to reach the source and drain depletion regions. The minority carriers then recombine with the returning majority carriers. The fast interface state component arises when surface states trap charge during inversion. As the surface returns to accumulation, these
charges are out of equilibrium; the charge stays trapped after the surface has returned to accumulation. These trapped charges then capture holes from the surface. This results in a net removal of charge from the source and drain regions.

The current contribution from the fast interface states is what is now identified as the charge pumping current. It can be calculated as follows.

\[ I_{cp} = qA_g(D_u)\delta V_s f \]  

(2.5)

The average density of interface states, \(D_u\), is multiplied by the swing in surface potential \(\delta V_s\), and the charge per trap \(q\) to get the areal charge. This, when multiplied by the gate area, \(A_g\), yields the total charge trapped per cycle. Finally, the total charge trapped per cycle is multiplied by the pulse frequency \(f\) to obtain the charge pumping current \(I_{cp}\). This equation may be used to extract the average interface density from charge pumping data.

Brugler and Jespers also used thermal stress bias experiments to alter the number of interface states. This provided further experimental support for the idea that interface states are responsible for charge pumping. Charge pumping current was measured for a matched pair of MOSFETs on a single chip. Next, they were stressed at a temperature of 400°C for a five minutes. The gate of one transistor was biased at \(-30V\); the gate of the other device was shorted. The high bias, high temperature condition creates additional interface states while the no bias, high temperature condition leaves the interface state density unchanged. After the stress, the charge pumping current was again measured. The device that was biased showed a higher charge pumping current. This shows that increasing the number of interface states increases the charge pumping current.
Charge pumping measurements on devices with several different devices and pulse waveforms provided experimental support for the geometric component model. Larger charge pumping currents were measured for long channel devices than were measured for short channel devices. Also, smaller charge pumping currents were measured when sawtooth pulses were used than when square pulses were used. Brugler and Jespers argued that the sawtooth pulses, which result in slower changes in gate voltage, allow more time for minority carriers to reach the source or drain region before majority carriers can migrate in from the substrate.

It should be noted, that this description is only partly correct. As later authors point out [43, 44, 45, 46], the essential characteristic of charge pumping is that the interface traps are out of equilibrium while the gate voltage is being swept. When sawtooth waveforms are used, the interface traps farthest from the center of the band gap are not out of equilibrium. Hence, they do not contribute to the charge pumping current. This feature of charge pumping currents was later exploited by Van den bosch, Groeseneken, Heremans, and Maes to extract interface trap density of states as a function of trap energy [47].

Later studies also point out that the geometric component can only be reduced to insignificance by using channels less than 5μm long [43, 45, 46]. Using hindsight, one can find evidence of this in Brugler and Jespers original paper. They measured the charge pumping current as a function of pulse amplitude for three different reverse biases. When the source and drain were not reverse biased, the charge pumping current did not saturate. In contrast, when the source and drain were reverse biased with respect to the substrate, the charge pumping current did saturate. However, the saturation current was a function of the bias voltage. These results may be understood
as follows. The source and drain depletion regions in the substrate are increase in size with increasing reverse bias voltage. This, in effect, reduces the channel length because the majority carriers from the substrate are expelled from depletion regions and the geometric component of the charge pumping current is less significant [48]. However, the threshold for charge pumping is increased. This is likely due the higher energy barriers that must be overcome to provide electrons for surface inversion from the source and drain (see Figure 2.7(a)).

2.2.5 Detailed Charge Pumping Model

A more complete understanding of the geometric component of charge pumping requires a more complete understanding of two things: minority and majority carrier movement and distribution during pulse transitions, and the non-equilibrium dynamics of interface traps. A more detailed and accurate description of charge pumping was first provided by Groeseneken in 1984, fifteen years after the discovery of charge pumping [43].

Consider the motion and distribution of charge during a single pulse cycle starting with an n-channel device in accumulation. As mentioned in the discussion of MOSFET operation, the source and drain are both reverse biased with respect to the substrate. The depletion regions do not have any free charge. The surface of the channel is in accumulation, so, there is an abundance of holes—the majority carriers—at the surface. During this part of the cycle, the gate channel Fermi level is below the intrinsic level and close to the valence band. Hence, most interface traps are filled with holes.
As the gate pulse moves from accumulation to depletion, the conduction and valence bands move down so that the Fermi level is closer to midgap. At flatband voltage, $V_f$, a depletion layer begins to form. Holes are driven from the surface. The free majority carriers respond very quickly, so they can stay in equilibrium [49]. If the transition is very slow, the interface traps remain in equilibrium. As the conduction band moves towards the Fermi level, the trapped holes are emitted into the valence band. These holes are swept out of the growing depletion layer into the substrate as free majority carriers.

If, however, the transition is rapid, the emission is not in steady state. The emission rate is slower than the change in surface potential. The majority carriers, holes, are rapidly swept out of the expanding depletion layer. This leaves a region with very little free charge. The traps cannot emit into the valance band. These hole-filled traps are now below the Fermi level.

As the surface moves past threshold, $V_f$, electrons sweep into the channel from the source and drain. Since the trapped holes are now below the Fermi level, they rapidly capture the electrons arriving from the source and drain. This removes net charge from the substrate. In the inversion regime, the surface is deeply depleted, and there are no charge carriers in the depletion region. The only charge carriers are the electrons at the surface [5, 48].

At this point, the Fermi level is close to the conduction band. Nearly all of the traps are rapidly flooded with electrons, and the traps return to equilibrium.

Now, consider the pulse from inversion to accumulation. As the gate voltage is reduced towards threshold, two things happen. First, the number density of electrons in the channel is reduced. Second, the traps emit electrons into the conduction band.
There, they are free minority carriers in the still inverted channel. Again, if the rate of surface potential change is slow enough, the traps remain in equilibrium and they empty into the conduction band. However, if the rate of surface potential change is faster, the traps are not in equilibrium. The emission rate is too low, and filled traps exist above the Fermi level. As threshold is crossed, they electrons start migrating towards the source and drain. The entire channel is nearly devoid of free charge.

As the gate voltage continues to drop below the flatband voltage, holes from the substrate rapidly flood and eliminate the depletion region. The traps filled with electrons are above the Fermi level. Hence, they rapidly emit the electrons into the valence band. This recombination results in a net removal of charge from the source and drain. This is the charge pumping current.

Now, the geometric component may also be explained by this model. As the gate voltage swings through the flatband voltage, holes rapidly flood the depletion region. The motion of the holes into the depletion region is faster than the motion of the electrons into the source and drain. The longer the channel is, the more holes reach the surface before the electrons from the inversion layer reach the source or drain [49]. Hence, these holes from the substrate will recombine with the electrons from the inversion layer. This results in the net removal of electrons from the source and drain and holes from the substrate. This current is in the same direction as the charge pumping current. However, unlike the current due to interface traps, it will not saturate as the top of the gate pulse is increased above inversion. Note also that the entire geometric component arises from the sweep from inversion to accumulation. In contrast, the trap component arises from both the inversion–accumulation and
accumulation-inversion transitions. In practice, the geometric component is negligible for any channels less than 5 \( \mu m \) in length [49].

This model of charge pumping lead to a method for calculating the density of states as a function of energy from charge pumping data [43]. Groeseneken noted that the rise and fall time of the gate pulse control which sub-population or interface traps contribute to the charge pumping current. The traps closest to the conduction or valence band edges have the fastest emission and capture time constants. Shockley-Read-Hall statistics may be used to calculate the emission time as a function of trap energy [40, 41]. Hence, from the rise and fall time of the pulse, the interface trap density at a given energy may be calculated. Combining measurements made with a variety of rise and fall times then provides a means of calculating the density of states. Further, varying the temperature can broaden the range of energies that may be probed with this technique [47].

2.2.6 Interface Studies with Charge Pumping

The theoretical explanation for charge pumping lead to experimental methods for studying interface traps. When compared to other methods, such as Deep Level Transient Spectroscopy (DLTS) charge pumping measurements have several advantages. First, unlike DLTS, charge pumping measurements can probe the density of states on both sides of the intrinsic level [43, 47]. Second, charge pumping measurements are made on technologically important MOSFETs. In contrast, DLTS measurements must be made on large area MOS capacitors [5, 47]. Further, direct comparisons of charge pumping measurements and DLTS show that charge pumping measurements are more sensitive and have better resolution [47, 50].
One use of charge pumping has been to study the role of hydrogen in altering defect densities [46, 51]. In the first set of experiments, MOSFETs were irradiated with 1 or 10 Mrad–(SiO₂). The devices were then placed either under vacuum or N₂ for several days. Next, the transistor was exposed to forming gas (1:9 H₂:N₂). During exposure, charge pumping measurements of the total number of interface states, Nᵢᵣ, the trapped positive charge, Nᵢₑ, and the density of interface states Dᵢᵣ were made. An increase in Nᵢᵣ and decrease in Nᵢₑ that were observed when irradiated devices were exposed to H₂. This data is explained by the following reaction sequence. H₂ reacts with a non-bridging oxygen atom to form H⁺. The H⁺ rapidly diffuses under the influence of the gate bias to the Si-SiO₂ interface and reacts with a Si-H bond to form H₂ and a dangling bond. The dangling bond is the new interface state.

Note that this data supports the ESR data which implicates dangling bond P₅ defects as interface states [16]. Further, the density of states data measured by the authors of this study show a change in Dᵢᵣ following the H₂ exposure that is similar in shape to that obtained from ESR measurements [51, 16]. This implies that the new interface states are P₅ centers, which are dangling bonds. Further evidence is that there is a strong dependance of Nᵢᵣ growth upon gate bias. When negative gate biases were applied, Nᵢᵣ did not increase.

A refinement of the charge pumping technique was used in a subsequent study of interface defect formation by atomic hydrogen [46]. In this modification of charge pumping calculations, numerical models of MOSFETs are used to include the contribution of the source and drain regions to the charge pumping current. This allowed a profile of the contribution to the charge pumping current as a function of position in the channel. In this study, MOSFETs were exposed to atomic hydrogen.
edges of the gate oxide at the source and drain regions were exposed to the ambient; a capping layer was not grown following metallization. Charge pumping measurements coupled with MOSFET simulations were then used to show that atomic hydrogen concentration decreases exponentially from the point of entry. The alteration in the density of states is, however, very similar to the alteration produced by H₂ on irradiated MOSFETs. This suggests that the surface chemistry is similar for the two treatments.

The sensitivity of charge pumping measurements is illustrated by a set of studies where individual interface traps were characterized with charge pumping [52, 53]. Submicron MOSFETs made with very low interface state densities may have only a single active trap. Yet, this trap can affect device operation. Hence, knowledge of the single trap is necessary. Using signal averaging techniques, charge pumping measurements with a resolution of ±1 fA were made. Standard charge pumping measurements were then used to measure the electron and hole capture cross sections for a single trap [53]. One knows that only a single trap was involved from the following argument. The maximum charge pumping current is given by

\[ I_{cp(max)} = qN_{it}f, \]

where \( q \) is the electronic charge, \( N_{it} \) is the total number of active traps, and \( f \) is the pulse frequency. A charge per cycle \( Q_{cp(max)} = I_{cp(max)}/f \) of \( 1.6 \pm 0.5 \times 10^{-19} \text{C} \) is then consistent with a single trap. However, it could still be partial contributions from multiple traps. This may be ruled out by increasing the peak to peak amplitude of the pulse. As the amplitude is increased, \( Q_{cp} \) increases if there is more than one trap present. The first trap they observed is shifted in voltage with respect to the second
trap by approximately \(-0.5\text{V}\). This is probably caused by the trap being located over the lightly doped drain. As described by Stahlbush [46], the local surface potential varies with position in the trap. Hence, the accumulation and inversion points vary with position, and the charge pumping curve is shifted.

Hole and electron capture cross sections were also measured for individual traps [53]. Here, the maximum charge pumped per cycle, \(Q_{it}\) is calculated from Shockley-Read-Hall statistics of interface traps [40, 41]. It is given by

\[
Q_{it} = qN_{it}(1 - e^{-t_f/\tau_{hp}}) \tag{2.7}
\]

\[
Q_{it} = qN_{it}(1 - e^{-t_r/\tau_{en}}) \tag{2.8}
\]

where \(q\) is the electronic charge, \(t_f\) is the fall time of the pulse, \(t_r\) is the rise time of the pulse, and \(\tau_{hp}\) and \(\tau_{en}\) are the hole and electron capture time constants. Note that in an n-channel device, the role of the rise and fall times is reversed. That is, electron capture processes occur on the falling edge of the pulse; hole capture processes occur on the rising edge of the pulse. The capture time constants are calculated with a simple kinetic argument. For example, a hole moving with a thermal velocity \(v_{th}\) through randomly distributed spheres of cross-section \(\sigma\) will collide with and be captured at a rate, \(r_{cp}\), given by

\[
r_{cp} = v_{th}\sigma p_s. \tag{2.9}
\]

where \(p_s\) is the surface density of holes. The time constant is simply the inverse of the rate. Similar equation gives the capture time constant for electrons. In addition.
the surface concentration of carriers can be readily determined from C-V and I_dV measurements. Hence, the capture cross section for holes and electrons can be calculated from Q_t measurements. Saks measured the capture cross sections for both holes and electrons as $\sim 10^{-16} \text{cm}^{-2}$.

**Charge Pumping Summary**

Charge pumping measurements provide a valuable tool for examining interface traps. This method allows the determination of the interface density of states function, the electron and hole capture and emission cross sections, and the spatial distribution of interface states in MOSFETs.

Charge pumping is an effect of non-equilibrium trap dynamics as the MOSFET gate is swung from accumulation to inversion and back. As the gate swings from accumulation to inversion, majority carriers are trapped. In inversion, the majority carriers recombine with minority carriers from the source and drain. As the gate then swings from inversion to accumulation, the minority carriers are trapped. In accumulation, they are emitted into the valence band where they recombine with majority carriers. This results in a DC current through the substrate.

Now, by taking advantage of the emission time energy dependance, one can vary the rate of gate swing and select the energy levels of the traps that contribute to the charge pumping current. In this manner, one builds a density of interface states function. Unlike DLTS measurements, this $D_{it}$ may be measured above and below midgap. In fact, it may be measured to within a few $kT$ of the band edge.

Examples of the use of charge pumping measurements include studies of the reaction chemistry at the Si/SiO$_2$ interface. It provides experimental support for a
cracking model of H$_2$ depassivation of interface states. Another example of the usefulness of charge pumping measurements is the determination of the properties of single interface traps.
2.3 BEEM compatible MOSFET Fabrication

Charge pumping experiments may be performed with any MOSFET that has an independent substrate connection. In contrast, BEEM measurements on a MOSFET structure require a MOSFET that has very thin gate metal (\(\sim 5 \text{ nm}\)) over a thin (< 10 nm) gate oxide. The thin metal gates are required so that tunnelling electrons from the STM tip are not scattered in the gate; likewise, the thin oxides are required so that tunnelling electrons that have crossed the gate are not trapped in the oxide (see section 2.2.2. Commercial MOSFETs do have very thin gate oxides. However, they also have thick, p-Si gates and are therefore not compatible with BEEM measurements. Early MOSFETs were made with metal gates but did not have thin oxides and so are not compatible with BEEM measurements [54, 55]. Therefore, a series of experiments were performed to evaluate potential fabrication processes for making a BEEM compatible MOSFET.

As a first step, simple MOS capacitors (MOSCAPs) were made by depositing a thin metal gate on a thin thermal oxide. BEEM and capacitance measurements were then made to evaluate the effect of photolithography processes on the gate oxide and gate metal. The results of these experiments indicated that photolithography processing would not alter the oxide or gate metal in a way that would prevent BEEM measurements.

Next, a simple MOSFET fabrication process based on the electrical engineering processing course (EE-637) laboratory manual was developed and tested [56]. This process included thermal oxidation, diffusion based dopant introduction, and thermal evaporation of Al. Particular attention was paid to passivating defects in the field oxides, making ohmic contact to the devices, and controlling gate oxide growth.
Devices made at OSU had a variety of interrelated defects. Contacts to the source drain region were either non-ohmic or the Al spiked through to the substrate. Diffusion junction depths were deep enough to prevent Al spiking or shallow enough to preserve a non-diffused region under the gate. Gate oxides were either thin enough for BEEM experiments or had the low interface state and fixed charge densities required for charge pumping experiments. Each of these defects may be separately eliminated with processes that require more sophisticated processing equipment than is available at OSU. So, arrangements were made to make several visits to the Stanford Nanofabrication Facility, a member of the National Nanofabrication Network. During these visits, a fabrication process that includes: semi-recessed local oxidation of silicon (LOCOS) for better field oxides, a field stop implant to pin the surface potential at the field oxide-Si interface, a backside gettering implant to immobilize substrate defects, ion implantation for better dopant distribution control, sputtering deposition of Al with 1%Si for ohmic contacts that do not spike through the diffusion layer, and low-pressure chemical vapor deposition (LPCVD) of poly-Si (p-Si) to protect the gate oxide during processing was developed. It is derived from the Stanford Nanofabrication Facility (SNF) BiCMOS process [57].

Much progress was made during the visits to the SNF. Good ohmic contacts to both the substrate and n⁺-implant regions were made. The combination of a semi-recessed LOCOS and field stop implants resulted in good isolation and pinning of the field oxide-Si interface. Substituting As implants for P diffusion yielded the necessary control over junction depth and lateral n⁺-implant region control. The initial oxide growth and anneal yielded good quality oxides. However, processing steps performed after the gate oxide was grown resulted in leaky gate oxides. Hence, a working
MOSFET compatible with BEEM measurements was not been made. The time and costs associated with working at the SNF forced us to seek another source of BEEM compatible MOSFETs. Fortunately, a collaboration with IMEC was developed. They are developing a process that is compatible with our requirements and have included devices designed to our specifications in their mask set. They will provide nearly finished devices: we will complete the gate metallization and perform the BEEM and charge pumping measurements.

2.3.1 Metal–Oxide–Si Capacitor Fabrication

The first step in evaluating the feasibility of fabricating a BEEM compatible MOSFET was to assure that minimal processing would not interfere with BEEM measurements. In previous BEEM work, Kaczer and Pelz [36, 38, 37] made MOSCAP structures by evaporating thin layers (~ 5 nm) of Pt or Pd through a shadow mask. In these experiments, the metal was not exposed to any chemical agents: in fact, most of the samples were made and tested in UHV conditions without breaking vacuum. In contrast, making a BEEM compatible MOSFET requires aligning the gate metal to the rest of the MOSFET components with micron or sub-micron accuracy.

This alignment accuracy generally requires photolithography and the chemical processing associated with it. So, we fabricated a MOSCAP with a lift-off metallization process (Figure 2.9). Our MOSCAPs were made on Si < 001 >, n-type, 0.004 – 0.02Ω-cm) wafers with 10 nm thermal oxides from National Semiconductor. The photoresist was applied and patterned with 250 x 250μm squares in a 1.1 mm pitch array by Dr. Richard Stahlbush [58]. A thermal evaporation was used to deposit ~ 5 nm of Pd on the pieces and the lift-off was performed by soaking and then
Figure 2.9: Schematic view of MOSCAP fabrication with a "lift-off" metallization process. (a) Photoresist is deposited on the SiO₂ layer and the top of the photoresist is hardened. (b) The photoresist is exposed, developed, and metal is evaporated onto the surface. (c) The photoresist is removed, taking the excess metal with it, and a MOSCAP remains.

spraying the samples with acetone until the excess metal was removed. The samples were then cleaned with acetone and methanol, and blown dry with N₂. They were then introduced into an ultra-high vacuum system (UHV) for capacitance-voltage (CV) and BEEM analysis. Annealing the sample at ~ 300°C and a background pressure of < 10⁻⁹T was found to be necessary for STM and BEEM measurements.

Two types of experiments were performed to evaluate the MOSCAPs. First, capacitance was measured as a function of gate voltage (CV-curves). This yields information about the electrical continuity of the metal square and the condition of the oxide. Second, STM and BEEM measurements were made to assess the effect of lift-off processing on BEEM measurements.
The CV-curves were obtained by applying a small AC voltage, at a variety of DC bias voltages, to the substrate and measuring the resulting AC currents. This procedure was repeated at a variety of bias voltages. The capacitance was calculated by using the relationship between capacitance $C$, applied AC voltage $V_{pp}$ at frequency $f$, and measured current $I_{pp}$:

$$C = \frac{I}{2\pi f V_{ac}}.$$  

(2.10)

Note that this equation is derived by assuming that there is no resistance in parallel or series with the capacitance. More sophisticated phase-sensitive measurements replaced this technique in later work.

A typical CV-curve obtained with a 50 mV peak-peak AC voltage at 100 Hz is presented in Figure 2.10. It is a normal, low-frequency CV curve. The capacitor is in accumulation at bias voltages above 3 V, in depletion between $-4$ and 3 V, and inversion below $-4$ V. These characteristics show that the oxide is not leaky. The effective area of the capacitor, calculated from the capacitance in accumulation, is $5 \times 10^{-4}$ cm$^2$. This is smaller than the measured area of the metal squares ($6.3 \times 10^{-4}$ cm$^2$). This may be the result of a discontinuous film, scratching of the film by the STM grounding wire.

BEEM measurements were successfully made on some of these MOSCAPs. Figure 2.11 is a schematic of the electronics used to collect BEEM data. The tip voltage $V_i$ is under computer control, the tunneling current is maintained by a feedback loop controlling the height of the tip above the sample, and the BEEM current $I_c$ is digitized and stored on the computer. The STM system and BEEM hardware have been previously described in detail [59, 15]. These images were obtained with the
Figure 2.10: Low-frequency $CV$ curve obtained from MOSCAP fabricated with a lift-off metallization. The metal square is 250 x 250$\mu$m. The AC frequency was 100 Hz.

Pd square grounded through a 100$\mu$m diameter Au wire. The tip (etched W) was brought into tunneling and a 4$nA$ tunneling current was maintained. For BEEM imaging, a suprathreshold bias, $-8V$ was applied between the tip and ground. The substrate was maintained at virtual ground by the current amplifier used to measure the BEEM current.

Figure 2.12 shows a scanning-tunneling microscopy (STM) topography image and the simultaneously acquired BEEM image of a 150 x 150 nm region on a MOSCAP. The grain structure of the Pd film is clearly seen in Figure 2.12(a). The small bright regions of Figure 2.12(b) show where there is BEEM current well above the noise background. BEEM images of MOSCAPs obtained by Kaczer and Pelz [36] show much better correlation with topography than that observed for MOSCAPs made
Figure 2.11: Schematic of the experimental setup used to acquire BEEM images. The tip current $I_t$ is kept constant by the feedback system that controls tip height above the surface. The tip voltage $V_t$ is under computer control and the BEEM current $I_c$ is digitized by and stored on a computer.

with lift-off metallization. This suggests that there may be chemical changes to the metal or contamination on the oxide under the metal. One method of differentiating between these possibilities is to examine BEEM-IV curves [36, 14].

As discussed in section 2.2.2, IV curves may be used to differentiate between regions with different energy barrier properties from regions with different transmission properties. Several example BEEM IV curves are presented in Figure 2.13. Each of the curves was obtained in a different region of a $150 \times 150\mu m$ scan. The Pd film was
Figure 2.12: Topography and BEEM images of a MOSCAP. The vertical scales are 1.5 nm for the topography image and 3 pA for the BEEM image. (data file: 9809236).

grounded. the substrate was held at virtual ground by the input to the current amplifier. and the tip voltage was scanned from -2 -10V. The color coding groups curves of similar threshold voltages. Note that all of the threshold voltages are greater than the ~ 4V threshold observed on Pd/SiO₂/Si by Kaczer and Pelz [36, 37, 38]. Note also. that each of the curves has a different amplitude and shape. This indicates that the transmission properties of the barrier as well as the barrier height vary across the sample. The conclusion is that the lift-off processing either altered the metal film-oxide interface properties or contaminated that interface.

Atomic Force Microscopy (AFM) and optical microscopy were used to examine the samples. Debris and other films were found on the sample. The likely source of the debris was incomplete cleaning of the samples following lift-off. This debris could be responsible for the difficulty in making BEEM measurements. In conclusion,
Figure 2.13: BEEM IV curves obtained at 6 different locations in a 150 x 150 nm region are shown. The color coding groups IV curves with similar threshold voltages. All of the IV curves shown have threshold voltages greater than the ~4V threshold voltage for SiO$_2$. (data file: 9809249).

BEEM measurements are possible on processed materials. However, further work is required to clean the surface following lift-off metallization to optimize the sample preparation for BEEM measurements.

2.4 MOSFET Fabrication at Ohio State University

The strategy used to design a fabrication process for BEEM compatible MOSFETs was to start with a simple process known to produce working transistors and make
minimal modifications to that process. We started with a diffusion based process used in the electrical engineering processing class. EE 637.

Simplification of device design was a further goal. Rather than making separate source and drain regions, which would require separate contacts in the UHV system, a rectangular channel inside a diffusion region was made. A top view of the device structure shows the alignment and scale of the various regions created during processing Figure 2.14(a). Cross-section schematics, drawn for the dashed line in (a), are presented in Figure 2.14(b) and (c).

Schematic cross sections through a MOSFET at each major step in the process are presented in Figure 2.15. In brief, a field oxide is grown and holes for the n⁺-doped regions (source, drain, test-structures) are etched through the oxide Figure 2.15(a). A two-step diffusion process—pre-deposition and a drive-in oxidation—forms the n⁺-doped regions and covers them with a protective oxide (Figure 2.15(b)). The gate regions are then exposed with a wet etch and the gate oxide is grown (Figure 2.15(c)). Next, contact vias are opened, lift-off lithography is done for the contact metal, the metal is deposited, and the excess metal is lifted off (Figure 2.15(d)). Finally, the gate metal lift-off is performed.

A more detailed outline flow chart for this process is presented in Figure 2.16. The first step in this process is thermal oxidation of Si wafers (p-type, 5-10Ω-cm, <001>) to grow a field oxide. The field oxide has several functions. During processing, it serves as a mask to define regions for dopant deposition, metal deposition, and gate oxidation. During device operation, field oxide serves to electrically isolate devices from one another.
The next step is to open windows in the field oxide, pre-deposit P dopant, and drive-in the dopant while oxidizing the Si in the doped regions. This is a two-step process. First, the wafers are exposed to phosphine gas (500 ppm in $N_2$) in an oxygen atmosphere at high temperature. This grows a phospho-silicate glass on the surface of the wafer and deposits P into the wafer at the solid solubility limit, $\sim 10^{20} \text{cm}^{-3}$. The glass is removed with a timed BHF etch that does not remove a significant thickness
Figure 2.15: Schematic cross-sections through a MOSFET following: (a) field oxidation, (b) diffusion, (c) gate area exposure and gate oxidation, (d) contact via etch and contact metallization, and (e) gate metallization.
Figure 2.16: (a) BEEM compatible MOSFET fabrication process flow chart
of the field oxide. Next, a drive-in oxidation grows a protective oxide layer over the diffused regions and completes the diffusion process.

With the source and drain regions formed, the gate region is exposed with a wet etch and the gate oxidation is performed. The gate oxidation is performed at lower temperatures than the other oxidations to better control the oxidation rate and uniformity. It is followed by a high temperature anneal that is critical to oxide performance. Without the anneal, oxides have high amounts of trapped charge, the source and drain diodes fail, and hence, the MOSFET fails (see section 2.1.1).

Next, contact vias are opened to the source and drain regions, test structures, and substrate with a wet etch. The contact areas are relatively large to reduce the contact resistance. Thermal or e-beam evaporation of Al is used for the contact metallization stage. Attempts were also made to use titanium-silicides/titanium nitride for the contact metal because of “spiking” problems, discussed in section 2.4.1. Contact metallization is followed by post-metallization anneal in N$_2$ at 400$^\circ$C for 20 min.

The final step is the gate metal lift-off. Metal deposition was done in a custom built thermal evaporator [15]. Typical gate metal films are 5 nm thick Pd. However, for test purposes, ~ 20 nm thick Pd films were also deposited. The lift-off was done by soaking in acetone, spraying with acetone, and cleaning with methanol and de-ionized water. Samples were blown dry with N$_2$.

The basic recipes for the fabrication steps are presented as a series of tables in Appendix A. For detailed, valve-by-valve instructions for OSU cleanroom procedures see the EE-637 Laboratory Manual [56].
2.4.1 OSU Fabrication and Device Evaluation

MOSFET fabrication is a series of steps which interact with each other in a complex manner. A problem in any one step can substantially impair device performance. Further, altering any one processing step can substantially impact the parameters required to make another step, either earlier or later in the process, function as required. Hence, it is essential to design a series of test devices and tests into any fabrication process. In the charge pumping device fabrication process we targeted testing at three critical areas: resistance of substrate and diffusion contacts; dose and distribution of dopants; and field, diffusion, and gate oxide integrity.

Two distinct approaches to testing were taken during the fabrication process: monitor wafers were used to evaluate the success of critical processing steps prior to continuing the fabrication sequence, and special structures were built into each sample for measuring the performance of device components. Monitor wafers were used for oxide thickness measurement, diffusion region surface resistivity measurement, etch rate evaluation for timing subsequent processing steps, and Si–oxide interface evaluation. Van der Pauw and Transmission Line Method (TLM) structures were incorporated for contact resistance measurements, and diffusion region resistivity measurements while MOSCAPs were included for tests of gate oxide integrity.

In addition, the health of diodes formed between the diffusion regions and the substrate was used to evaluate dopant dose and distribution, contact resistance, and field oxide–Si interface properties. In the final analysis, the wealth of information available from MOSCAP and diode measurements provided the most useful assessments of fabrication process design. A set of charge pumping devices, MOSCAPs, TLM structures, Van der Pauw structures, and substrate contacts was included in each 5 x 20
mm area, which I will call a chip. There were 8 chips included on each wafer. This provided ample spacing between chips for dicing the wafer by cleaving the wafer.

The simplest component, the substrate contact, plays an important role in the performance of all devices on the wafer. Fortunately, it is also the easiest to fabricate because the wafers are p-type Si and Al is a p-type dopant that readily diffuses into Si [56, 55]. Multiple substrate contacts were included in each chip to provide redundancy and a means of testing the contacts. The simplest test was usually sufficient. IV measurements between two substrate contacts were made. In general, the contacts were ohmic and the contact resistance was estimated. Four point probes were not used because the resistance between two probes on the same pad was < 1Ω. A typical substrate–substrate contact IV curve is presented in Figure 2.17. The resistance of the substrate ($\rho = 5 \times 10 \Omega \text{cm}^2$) is ~ 1kΩ so the contact resistance is negligible.

The next test is to evaluate the behavior of the diodes formed between the diffusion regions and the substrate. Simple IV measurements yield much information about the health of a diode so they were done first. Figure 2.18 shows forward and reverse biased IV-curves for a single diode. It is clear that these are not good diodes: the IV should be linear for forward bias when plotted on a semilog plot, the current should be higher, and there should not be the large leakage current observed.

Two principle causes have been identified for the defects observed in these diodes. First, an inversion layer is present at both gate and field Si/SiO₂ interfaces. This inversion layer forms surface channels which vary the effective area of the diode as a function of bias [39]. Further evidence for inversion regions under the field oxide is presented with the MOSCAP measurement discussion. Second, the reverse bias leakage current is very large. This is caused by Al “spiking” through the pn-junction.
Capacitance voltage characteristics can provide a great deal of information about the success or failure of a fabrication process. Example CV measurements made on a MOSCAP produced with the OSU fabrication process are shown in figure 2.19. Both the low and high frequency CV curves indicate that there is fixed charge at the interface. The expected low-frequency flatband capacitance for this capacitor is \( \sim 100 \) pF so the flatband voltage is \( \sim -0.3 \) V. The difference in work function for p-type Si (\( 2 \times 10^{15} \text{cm}^{-3} \)) and Al is \(-0.9 \) V [4]. The difference between this and the flatband voltage is caused by charge in the oxide. The charge number density, assuming that all the fixed charge is concentrated at the interface, is \( 10^{11} \text{cm}^{-2} \). Note
Figure 2.18: Forward (a) and reverse bias (b) IV curves for the pn junction formed by the charge pumping device diffusion region and the substrate. Note the non-ideal behavior (ideality factor $\sim 2.5$) and the high reverse bias leakage current. See the text for a discussion of the causes of failure in these diodes.

that this is the sum of all charge sources. Individual contributors such as fixed oxide charge (generally negative) and mobile Na ions (positive) may well have number densities hundreds or thousands of times higher.

The high frequency CV characteristic provides a clue that there is fixed negative charge at the Si/SiO$_2$ interface under the field oxide. There is an increase in capacitance at gate voltages that would invert the Si surface if the minority carriers could respond. This feature is likely caused by an inversion layer under the field oxide that provides a source of minority carriers to the surface under the capacitor being tested [5].

Al contacts fabricated at OSU showed good ohmic properties. However, the Al contacts to diffusion regions spiked through and shorted the pn–junctions. Figure 2.20 is two scanning electron micrographs (SEMs) of an OSU charge pumping device that
has had the Al contacts removed by a wet etch. The sample was then sputter coated with gold for SEM imaging. Note that the edges of the contact regions have deep depressions that are very typical of Al spiking [55].

The problems that were encountered in making BEEM compatible MOSFETs required that we significantly increase the complexity of our fabrication process. Oxide charge and the resulting surface inversion, a problem common on p-type substrates [5], required better device isolation. A semi-recessed local oxidation of Si (LOCOS) with field stop implants was implemented to control field oxide charge [4]. Spiking of Al through the diffusion region shorted the source and drain regions to the substrate. Replacing the thermal or e-beam evaporated Al with sputtering of an Al/Si alloy to form the contacts was chosen as the solution to this problem [55]. Finally, the ultimate quality of the gate oxide was marginal. Better control over oxidation and anneal time and temperature were available at SNF and used to improve oxide growth.
2.4.2 MOSFET Fabrication at Stanford Nanofabrication Facility

An overview of the process developed at the Stanford Nanofabrication Facility (SNF) to overcome the defects observed in the OSU fabrication process is presented in Figure 2.21. The left side of the figure shows schematic top views of a MOSFET at various stages of the process and the right side of the figure shows schematic cross-sections of a MOSFET at the same stages of the process. The cross-sections are taken along the dashed lines on the top view. The first two major steps in the SNF fabrication process is the LOCOS with field stop implant. Figure 2.21(a) shows: the pad oxide, the silicon nitride oxidation mask, and the B field stop implant. Then the field oxidation is done (Figure 2.21(b)). Next, the gate oxide is grown and the As source/drain implant is performed; photoresist is used as the mask (Figure 2.21(c)). The next step is to open contact vias and deposit the Al/Si alloy contact metal (Figure 2.21(d)). Finally, a lift-off process and a thermal deposition is used to deposit
the gate metal (Figure 2.21(e)). Complete descriptions of the recipes used in the SNF fabrication process are presented in Appendix B.

The MOSFET fabrication process developed at SNF starts with a semi recessed LOCOS process. A flow diagram for this step is presented in Figure 2.22. First, a thin thermal oxide pad (~25 nm) was grown on the wafers to relieve stress in the subsequently deposited nitride film. The wafers are immediately transferred from the oxidation furnace to the low pressure chemical vapor deposition (LPCVD) furnace and a ~85 nm layer of Si$_3$N$_4$ was deposited. Photolithography was used to define the regions where devices were to be fabricated and a plasma etch was used to remove the Si$_3$N$_4$ over field oxide regions. Following the field stop implant of B (Table 2.1), the devices are at the stage depicted in Figure 2.21(a). The nitride pads acted as an oxidation mask because oxygen does not readily diffuse through Si$_3$N$_4$. Next, ~650 nm of field oxide was grown. A "bird's beak" formed at the edges of the nitride pad because oxygen diffused through the pad oxide and oxidized the Si under the edges of the nitride pad [4, 55]. This deformed the edges of the nitride pad as illustrated in the cross-section of Figure 2.21(b). The final step of the LOCOS process is to remove the nitride pad. The nitride pad removal is a four step process: the nitride was stripped in a hot H$_3$PO$_4$ acid bath, the pad oxide was removed with a buffered HF (BHF) dip, a sacrificial clean-up oxide was grown, and the sacrificial oxide was removed with BHF. Finally, the B field stop implant was performed.

The next steps in MOSFET fabrication are gate oxidation, source/drain implantation, a backside gettering implant, and a dopant activation anneal. These processing steps are summarized in Figure 2.23. The gate oxide was grown at low temperature (850°C)—to provide a slow oxide growth rate—in a furnace that had been cleaned
Figure 2.21: Schematic view of the MOSFET fabrication process used at the SNF. (a) and (b) show views of a MOSFET during and following LOCOS. (c) shows the gate oxide and source/drain As implant. (d) shows the source/drain contact formation and (e) shows the completed MOSFET.
Figure 2.22: Flow diagram for the semi-recessed local oxidation of Si process used at the SNF.
Table 2.1: Implant Parameters. The source/drain implant was done by Nu Ions of San Jose, CA. All others were done at the SNF.

<table>
<thead>
<tr>
<th>Implant Function</th>
<th>Ion Species(s)</th>
<th>Ion Energy (keV)</th>
<th>Dose (cm⁻²)</th>
<th>Wafer Numbers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Channel Stop</td>
<td>¹¹B⁺⁺</td>
<td>99.2</td>
<td>1x10¹³</td>
<td>30.32-36.38.39</td>
</tr>
<tr>
<td>Channel Stop</td>
<td>¹¹B⁺⁺</td>
<td>100.0</td>
<td>1x10¹³</td>
<td>41-49</td>
</tr>
<tr>
<td>Source/Drain</td>
<td>⁷⁵As⁺⁺</td>
<td>80.0</td>
<td>1x10¹⁶</td>
<td>32-36.38.39.41-49</td>
</tr>
<tr>
<td>Backside Gettering</td>
<td>⁴⁰Ar⁺⁺</td>
<td>100.0</td>
<td>3x10¹⁵</td>
<td>32-36.38.39.41-49</td>
</tr>
</tbody>
</table>

overnight with TCA. The TCA clean of the furnace is a chlorine based clean that scavanges sodium ions from the furnace and prevents the incorporation of Na ions into the oxide. After gate oxidation, photolithography was used to mask all parts of the wafers where As implants were not to go and the As implant was done to create the source and drain regions of the MOSFETs. Note that a plasma ashing was required to remove the photoresist following the implant because the implants had hardened the resist too much for chemical (piranha etch) removal to be effective. Then a backside gettering implant was done. This implant creates defects in the substrate which scavenge impurities during the dopant activation anneal [4]. The dopant activation anneal serves three purposes: anneal the gate oxide to reduce the fixed oxide charge [5], scavenge substrate impurities by “activating” the backside gettering implant, and recrystallize the Si damaged by the source/drain implant [4]. Following the anneal, the devices are at the stage depicted in Figure 2.21(c).

Contact formation to the source, drain, and gates are the final steps in the SNF MOSFET fabrication process and are summarized in Figure 2.24. First, the gate...
Figure 2.23: Flow diagram for gate oxidation, source/drain implants, backside gettering implant, and dopant activation anneal
oxides must be protected from the plasma etch used to assure that there is no oxide in
the bottom of the contact vias so \( \sim 100 \) nm of p-Si was deposited with LPCVD. Next, photolithography was used to define the contact vias and plasma etching was used to open the contact vias. After cleaning, the wafers are introduced into a sputtering chamber, a plasma etch removes any native oxide in the contact vias, and 500 nm of Al/(1% Si) film is sputtered onto the wafers. The Si in the Al prevents spike formation during the post-metallization anneal. The metal is patterned with photolithography and the excess metal is removed. A chemical etch (see Appendix A) was used on some wafers and dry etching (P-5000 etcher) was used on the remaining wafers. A forming gas \((\text{N}_2/\text{H}_2)\) post-oxidation anneal was used to complete source and drain contact formation. Finally, the wafers were diced and a lift-off process was used to deposit 5 nm of Pd for the gate metal.

A photo of a finished MOSFET is shown in Figure 2.25. There are several features that are implemented to allow access to the device in the UHV-STM chamber. The gate contact pads are separated from the device to allow room for the STM tip in the gate channel region of the device after landing a Au wire (100\(\mu\text{m}\) diameter) on the gate contact pad. Similarly, the contacts to the source and drain regions are connected via contact wires (not shown for this device) to contact pads that are at the edge of the sample. This allows source and drain contacts to be made without interfering with STM access to the MOSFET. Also visible in this image are alignment errors introduced during photolithography. The errors in this image are approximately 4 \(\mu\text{m}\) in the vertical direction and 0.6 \(\mu\text{m}\) in the horizontal direction. This wafer had the largest alignment errors of the wafers produced at the SNF. Most alignment errors were less \(\sim 1\mu\text{m}\).
Figure 2.24: Flow diagram for the source/drain contact formation and gate contact lift-off performed at the SNF.
Figure 2.25: Picture of a completed MOSFET. The probe damage was inflicted during electrical testing of this device. The vertical alignment error between contact metal and the active region was 4 μm; the horizontal alignment error was \( \sim 0.5\mu m \).

2.4.3 OSU Fabrication and Device Evaluation

Testing of the devices fabricated at the SNF revealed that many of the problems encountered at OSU had been solved. The contacts to the source, drain, and substrate regions were all found to be ohmic and low resistance. An example IV characteristic curve between two substrate contacts is presented in figure 2.26.

In contrast to the devices fabricated at OSU, typical diodes made at the SNF showed good to excellent reverse bias current characteristics and reasonable forward bias performance. The reverse bias currents for the SNF diodes that were tested ranged from pA to hundreds of pA at 4 V. An example forward biased IV characteristic is shown in figure 2.27. It shows higher currents than those obtained on diodes fabricated with the OSU process. This indicates that the contact problems and field
Figure 2.26: An IV characteristic between two substrate contacts shows low resistance (263Ω) and excellent linearity. The dots are data points and the red line is a linear least squares fit to the data (Pearson's r = 0.999).

oxide inversion problems observed with OSU devices have largely been solved with the SNF process.

While the devices fabricated at SNF are better than those fabricated at OSU, they are still not working MOSFETs. Time and funding constraints forced us to stop trying to fabricate the devices ourselves. Instead, the Pelz group is currently pursuing these experiments via a collaboration with IMEC in Leuven, Belgium. They will provide working, tested MOSFETs and the Pelz group will modify them to make them BEEM compatible and then perform BEEM and charge pumping experiments.
Figure 2.27: A forward bias IV characteristic for an SNF diode shows good performance. The maximum reverse bias current observed (at -4 V) was 8 pA.
CHAPTER 3

SCANNING CAPACITANCE MICROSCOPY FOR THIN FILM CHARACTERIZATION

3.1 Introduction

Dielectric thin films play an important role in semiconductor, hard disk drive, and Micro electro-mechanical systems (MEMS) technologies [7, 4]. In these applications, dielectric film thickness variations on nm-lateral scales can cause changes in device operation. For example, the MEMS motor pictured in figure 1.1 is susceptible to high stiction, friction, and early failure caused by “holes” in the lubricant films [6]. However, traditional methods of determining film thickness —such as ellipsometry, x-ray photoemission spectroscopy, and capacitance metrology—measure average thickness with lateral resolution of at best ~ 10 μm. In contrast, developing new applications of thin film technology will require nm-scale lateral measurements of dielectric film thickness variations.

This chapter addresses a fundamental metrology problem. How can the thickness of technologically important molecularly thick lubricant films be measured with sub-nm vertical resolution and sub-micron lateral resolution? Conventional methods of measuring the thickness of nm-scale thick films have sub-nm vertical resolution
but rely on spatial averaging so they have lateral resolution of tens to thousands of microns [8]. In this work, a scanning capacitance microscopy technique was developed to improve the lateral resolution of thickness measurements.

The idea behind SCM for thickness measurements is simple. Any two electrical conductors separated by an insulator form a capacitor. The capacitance of the system depends on the thickness and dielectric constant of the insulating layer. In scanning capacitance microscopy, an AFM tip is one of the conductors, the substrate is the other conductor, and the dielectric film is the insulator. As the AFM tip is scanned across regions of the dielectric film with different thicknesses, the capacitance should change. This work was directed at measuring that capacitance and quantitatively relating the measured capacitance changes to thickness variations.

Section 3.2 in this chapter is a description of the instrumentation developed for direct, low-frequency scanning capacitance microscopy is described and characterized. Additionally, the complications due to parasitic capacitances are described and for the first time, quantified.

Section 3.3 in this chapter is a description of the process developed to make thin film thickness measurements with scanning capacitance microscopy. Ultimately, we were able to make nm scale thickness measurements with ~ 200 nm lateral resolution.

3.2 Instrumentation for direct, low frequency scanning capacitance microscopy

In scanning capacitance microscopy (SCM), a nanometer scale conducting tip is scanned across the sample surface and a capacitance detector measures variations in the probe-sample capacitance \( C \). To date, the most common applications of SCM have been semiconductor characterization including dopant profiling [60, 61], device
characterization [62, 63], and surface defect characterization [64]. The common thread in these measurements is that the samples exhibit voltage dependent capacitance due to a voltage-dependent space charge layer in the semiconductor substrate. These implementations of SCM generally do not measure $C$ directly. Rather, they measure $dC/dV$ by varying the probe-sample voltage $V$ at frequencies greater than 10 kHz [60, 61].

However, not all systems of interest contain a semiconductor space charge layer that exhibit voltage dependent capacitance. This is particularly true for thin dielectric films on highly conducting (metallic) substrates. In this case, it is necessary to measure the capacitance directly rather than measuring $dC/dV$. Examples of such thin-film systems include perfluoropolyether compounds that are of fundamental importance in lubrication of disk drive and micro-electromechanical machine systems (MEMS) [7]. Previously, a macroscopic form of a “direct” SCM has been used to monitor thin, perfluoropolyether lubricant films with mm-scale lateral resolution by applying an AC voltage and measuring the resulting displacement current [8]. The original motivation for this present work is to extend such a direct SCM technique to a lateral resolution of <500 nm.

This paper addresses two main topics. First, we describe a simple approach, based on commercially available electronics, that can measure the capacitance with sub-aF/$\sqrt{\text{Hz}}$ noise levels at frequencies down to 1 Hz. Our approach can also distinguish between displacement and dissipative currents. In essence, this is a combination of two techniques, Scanning Capacitance Microscopy and Scanning Spreading Resistance Microscopy. We also characterize the noise in this system.
Second, we present detailed measurement and analysis of how the long-range stray capacitance between the sample and the mm-sized probe assembly varies with the lateral and vertical probe position. It turns out that these variations can be quite large (> 500 aF) and must be recognized, characterized, and subtracted from the measured capacitance signal in order to quantify small nm-scale lateral variations in the capacitance between the sample and the nm-scale probe tip.

### 3.2.1 Scanning Capacitance and Resistance Measurements

A drawing of an AFM cantilever and tip while engaged on the surface is presented in figure 3.1(a). The emphasis here is that the capacitance between the AFM probe and sample includes the capacitance of the nm-scale probe tip, the μm-scale cantilever, the mm-scale cantilever chip, the chip holder, and the associated wiring. The equivalent circuit is shown in figure 3.1(b). The tip-sample capacitance $C_{ts}$ is typically expected to be only about $10^{-40}$ aF for a 50–150 nm radius tip[65]. In contrast, the stray capacitance $C_{stray}$ is on the order of $1/2$ pF. Further, there may be a finite resistance $R_{film}$ between the tip and the substrate.

The principal challenge in SCM is measuring tip-sample capacitances that are on the order of an aF in the presence of stray capacitances that are on the order of $1/2$ pF [66, 67, 65]. Any detection technique must be capable of either measuring capacitance to a part per million or compensate for the effect of the stray capacitances. In addition, the accurate nm-scale analysis of thin dielectric films requires the detection and measurement of any finite resistance at the same time that one measures small variations in $C_{ts}$.
The most common means of compensating for the effect of the stray capacitance in SCM is to detect capacitance changes with ultra-high frequency tuned circuits (typically ~ 900 MHz). The RCA tuned circuit is an example of this class of detector (figure 3.2(a)) [68, 69]. In this type of circuit, the stray capacitance is included in the tuning elements. Hence, the small changes in capacitance caused by tip-sample capacitance variations change the resonant frequency of the circuit. When the circuit is tuned just off resonance, these small capacitance changes result in large changes in signal amplitude at the output of the circuit. This variation in amplitude is detected with a diode demodulator. Under ideal conditions, this circuit can be sensitive to ~ $10^{-21} \text{F}/\sqrt{\text{Hz}}$ changes in capacitance at modulation frequencies above ~ 10 kHz [66, 70]). However, it has several disadvantages. First, it is difficult to make quantitative capacitance measurements because the sensitivity is altered by any changes in stray capacitance. Second, the tuned circuits noise spectra are dominated
Figure 3.2: (a) Schematic of a typical RC type tuned circuit used in commercial SCMs. A UHF signal is inductively coupled to a tuned circuit that includes the AFM tip-sample capacitance. The tuned circuit is inductively coupled to a diode demodulator detector circuit. (b) Schematic of a bridge circuit for detecting small changes in capacitance against a large background capacitance $C_p$. The nulling capacitor is varied until there is no output voltage. Variations in $C_{ts}$ then unbalance the bridge and create an output voltage. A lock-in amplifier is typically used to amplify the signal and reject the noise.

by $1/f$ noise at frequencies below $\sim 5$ kHz [71, 66]. This makes direct measurements of $C$ difficult. Finally, the tuned circuits cannot detect the existence of resistive contacts: the presence of a resistive contact simply makes the circuit less sensitive [66, 69]. We also point out that commercially-available SCM's are much noisier than the optimized $10^{-21}$aF/$\sqrt{\text{Hz}}$ level. For example, Digital Instruments has reported a noise level of $4-11$aF/$\sqrt{\text{Hz}}$ (at 90 kHz) for their SCM sensor [72].

Bridge circuits have also been used to compensate for the effect of the stray capacitances on SCM measurements made with scanning tunneling microscopy (STM).
The currents flowing through the two arms of the bridge are equalized by changing the variable capacitor until it matches the stray capacitances. When balanced, there is no voltage difference between the two arms of the bridge. Changes in tip-sample capacitance then alter the current balance and a voltage difference appears across the bridge test points. The sensitivity of the commercial bridge circuits used was reported to be about 2 aF[73]. However, this sensitivity comes at the price of a narrow bandwidth.

For this paper, we have employed a third method, which is similar in principal, but different in implementation, to a bridge circuit. It is based upon a technique used by Pelz and Koch [75] to cancel stray capacitive currents in scanning tunneling microscopy potentiometry experiments. Our circuit is shown in figure 3.3. It uses phase-sensitive detection to measure the current that flows when a fairly low frequency (\( \sim 5 \, \text{kHz} \)) sinusoidal voltage is applied between the sample and the tip. The left arm of the circuit includes the tip-sample capacitance \( C_{ts} \), the stray capacitance \( C_{stray} \), and the film resistance \( R_{film} \) (figure 3.1(b)). The right arm includes a "bucking circuit", described in Appendix A, which is used to add an inverted, constant current that cancels the displacement current flowing through \( C_{stray} \). This bucking circuit consists of a variable gain amplifier, a variable phase shifter, and a coupling capacitor, \( C_{ce} \). The net current from the two arms of the circuit is measured with a sensitive current amplifier. The displacement (capacitive) and dissipative (resistive) components of the current are measured with a two channel lock-in amplifier.

The variable gain amplifier and variable phase shifting circuit are used for the following reason. When a sinusoidal drive signal of amplitude \( V_0 \) and frequency \( \omega \) is applied to the sample, a current \( I_{afm} \) is driven through the probe-sample junction,
and a bucking current $I_{bc}$ is driven through the coupling capacitor $C_{cc}$. The currents are given by

$$I_{afm} = j\omega V_0 (C_{ts} + C_{stray}) + \frac{V_0}{R_{film}}. \quad (3.1)$$

and

$$I_{bc} = -j\omega GV_0 C_{cc}. \quad (3.2)$$

where $G$ is the bucking circuit gain. This assumes that the phase of the bucking current is adjusted so that it is $180^\circ$ out of phase with respect to the signal applied.
to the sample. When \( G \) is adjusted so that \( G = C_{stray}/C_{cc} \), the net current at the input to the current amplifier is given by

\[
I_{net} = j\omega V_0 C_{ts} + \frac{V_0}{R_{film}} \equiv jI_{im} + I_{real}. \tag{3.3}
\]

The displacement current through the large stray capacitance has been cancelled by the bucking circuit so the current amplifier may be operated at high gain without being overloaded. The in-phase \((I_{real})\) and out-of-phase \((I_{im})\) parts of \( I_{net} \) are measured independently with a dual-channel lock-in amplifier as the tip is scanned.

From these phase sensitive measurements of the current, the capacitance change, \( C' \), and parallel resistance variation, \( R_{film} \), may be calculated by rearranging and separating equation 3.3 into its components as follows: \( C' = I_{im}/(2\pi f V_0) \) and \( R = V_0/I_{real} \), where the amplitude of the applied voltage is \( V_0 \), the displacement current is \( I_{im} \), the dissipative current is \( I_{real} \), and the frequency of the applied voltage is \( f \). Note that the measured capacitance \( C' \) is offset from the total capacitance by an amount determined by the bucking circuit gain. The current \( I_{net} \) was measured with an Ithaco 1212 current amplifier operated at the \( 10^{-8} \) A/V range setting.

The AC voltage frequency (5 kHz) and amplitude (3.7 Vrms) were selected to maximize the signal to noise for our equipment. Frequency was limited by the high frequency roll-off of the Ithaco current amplifier. In most of our measurements, the lock-in amplifier output filter (\( \tau = 1 \) ms) limited the measurement bandwidth to 160 Hz.
Sensitivity and Noise Analysis

We have measured the noise present in our capacitance measurements with both a test circuit and an AFM probe assembly mounted on a Digital Instruments Dimension 3000 AFM connected to the measurement circuitry. While the circuitry was attached to the AFM, noise measurements were made with either: (1) the tip engaged on the surface but not scanning, (2) the tip engaged and scanning, (3) or the tip retracted from the surface by several microns. The test circuit is composed of a capacitor ($C_{test} \sim 0.5\text{pF}$) connected between the AC voltage source and the current amplifier to simulate the probe-sample capacitance and a second capacitor connected between the bucking circuit and the current amplifier to simulate the coupling capacitor $C_c$. Comparing these noise spectra allows us to differentiate between noise induced by the measurement circuitry, noise that is inherent to AFM, and noise induced by scanning the AFM tip across a sample. Typical noise spectra are presented in figure 3.1. These spectra were obtained by connecting the capacitive channel analog output of the lock-in amplifier to the input of a Stanford Research Systems SR-770 spectrum analyzer. A series of 20-25 spectra in the frequency range from DC to 195 Hz were obtained and averaged.

The spectra presented in figure 3.4(a) were obtained with the test circuit connected to the system. The lower spectrum was measured with the test circuit connected directly to the current amplifier. The spectrum is very flat and the average noise level is $\sim 0.25\text{aF/}\sqrt{\text{Hz}}$ all the way down to 1 Hz. The upper spectrum in figure 3.4(a) was measured with an additional 30 cm of coaxial cable between the test circuit and current amplifier. This cable is comparable to the coaxial cable connecting the current amplifier to the AFM tip holder. This extra coaxial cable raises the base noise level
Figure 3.4: (a) Noise spectra obtained with circuit connected directly to a fixed capacitor test circuit (solid line), and with ~ 30 cm coaxial cable inserted between capacitor and current amplifier (dashed line). (b) Noise spectra obtained with measurement circuit connected to the tip. The tip was on the sample surface and not scanning (solid line) and on the surface during a 10µm scan (dashed line). The non-scanning noise level is < 0.4 aF, which is similar to that measured with the test circuit.

to 0.35aF/√Hz but does not add 1/f noise. While optimal RCA-type circuits can have noise levels as low as 10^{-21} aF/√Hz at frequencies above 4-5 kHz [71, 66], they exhibit significant 1/f noise levels at measurement frequencies closer to 1 Hz. The noise spectra from the resistive channel of the lock-in amplifier (not shown) are also flat down to 1 Hz, with a magnitude of less than 35 fA/√Hz. This is comparable to the noise currents in the capacitive channel of the lock-in amplifier. This current noise is mostly due to the noise current from the Ithaco current amplifier which we measured as 30 fA/√Hz at 5 kHz.

The spectra presented in figure 3.4(b) were obtained with the measurement circuitry connected to an AFM probe. When the tip is not being scanned, the low
frequency noise behavior matches that of the test circuit connected to the current amplifier with a ~ 30 cm length of coaxial cable. In this case, there is little $1/f$ noise down to 1 Hz and the base noise level is $0.35 \text{aF}/\sqrt{\text{Hz}}$. However, these spectra do contain spikes at frequencies from 20 to 200 Hz that are probably caused by electrical or vibrational pick-up in the laboratory. These peaks are not large and add little integrated noise power. The noise level integrated over a 160 Hz bandwidth, measured with the tip not scanning, is 4.1 aF. In comparison, with the test circuit connected with the 30 cm coaxial cable, the noise level integrated over a 160 Hz bandwidth is 4.0 aF. Now, when the tip is scanning over a 10μm lateral scan (dashed line in figure 3.4b), the spectrum does show significant power at frequencies below 2 Hz. This is caused by position-dependent changes in the measured capacitance, which are discussed in section 3.2.2. Scanning does not add significant noise at frequencies above 2 Hz.

In addition to the noise described above, there is a slow drift in the output of the capacitance measurement circuit. When connected to the test circuit, the drift is ~ 4 aF/min. We note that a drift rate of $40 \mu \text{V}_{\text{rms}}/\text{min}$ (approximately 1 part in $10^5$ per minute) in the output of the bucking circuit would cause the observed drift. While connected to an AFM tip in contact with a sample, the drift rate increases to 6–9 aF/min. Note that this drift does not add observable noise to the measured spectrum above a frequency of 1 Hz.

In summary, our measurement circuitry exhibits ~ $0.35 \text{aF}/\sqrt{\text{Hz}}$ of noise from 1 to 160 Hz, with an integrated noise level of < 5 aF over this bandwidth. This noise is mostly due to current noise from the current amplifier. This suggests that using
a low noise, higher frequency current amplifier and a higher frequency sinusoidal AC voltage could increase sensitivity further still.

3.2.2 Changes in Stray Capacitance Due to Scanning

As discussed in section 3.2, the goal of SCM is to measure small variations in the nm-scale capacitance $C_{ts}$ associated with the nm-scale probe tip and the local sample region. However, this must be done in the presence of a large (0.5 pF) stray capacitance $C_{stray}$ that is due to long-range capacitive coupling between the sample and the $\mu$m-scale cantilever and mm-scale cantilever support structure. We have found that this long range $C_{stray}$ itself exhibits significant variations as the probe is scanned over a sample. Hence, in order to properly monitor and quantify nm-scale variations in $C_{ts}$, it is necessary to characterize and understand how $C_{stray}$ varies with tip position.

This is illustrated in figure 3.5, which shows simultaneously acquired topography and capacitance data obtained while scanning across a “trench”. The sample (calibration grating TGZ02 supplied by MikroMasch) is a grating of trenches etched in SiO$_2$ on Si and coated with 10 nm of Si$_3$N$_4$. Each trench is specified as 104 ± 1.5 nm deep and 3 $\mu$m wide. figure 3.5(a) and (c) are grey-scale images of the topography and capacitance, respectively. figure 3.5(b) and (d) are cross sections along the dashed lines in figure 3.5(a) and (c). There are two puzzling features in the capacitance cross section. First, the SCM cross section is varying in the flat regions away from the step. This variation is nearly linear with a slope of ~100 aF/$\mu$m (figure 3.5(d) dashed line). One would naively expect that the nm-scale tip-sample capacitance $C_{ts}$ should be essentially constant in these flat, uniform regions. Second, there is a large change
in capacitance ($\sim 350$ aF) when the probe crosses the edge of a trench. From models of the tip-sample capacitance, one would expect that $C_{is}$ variations should only be $10s$ of aF when the probe is scanned across this trench boundary [65].

We have found that these variations in measured capacitance are due to variations in the long-range capacitance $C_{stray}$. $C_{stray}$ varies with tip position in two principal ways. First, $C_{stray}$ varies systematically with lateral tip position, even when scanning a uniform sample that has no intrinsic variations in dielectric film thickness or morphology. These lateral variations can range from $\sim 1$ to $\sim 100$ aF/μm of lateral scanning. The largest variations occur when the long axis of the cantilever is parallel to the fast scan direction while smaller variations occur when the long axis of the
cantilever is perpendicular to the fast scan direction. We attribute these lateral variations to changes in the "tilt" of the probe assembly with respect to the sample that are caused by the bending of the piezoelectric scanner tube. Second, $C_{stray}$ varies with the average height $h$ of the probe assembly above the sample surface. Typical capacitance variations with height are $\sim 2-4$ aF/nm, depending on cantilever size and shape. This is simply due to the change in separation between the sample and the probe assembly. These two factors can explain the anomalous features in the capacitance trace shown in figure 3.5(d).

**Stray Capacitance changes with lateral probe position**

To characterize variations in $C_{stray}$ with lateral position, we scanned a flat, uniform, 10 nm thick, thermal SiO$_2$ film grown on highly doped Si. On such a uniform sample, we would not expect any variations in the nm-scale capacitance $C_{ts}$. In fact, we observed a nearly linear variation across the sample, in the total capacitance $C_{ts} + C_{stray}$. Figure 3.6(a) is a graph of two single scan lines to illustrate this effect. When the long axis of the cantilever is parallel to the fast scan direction (which we call a "0° scan," solid line) the slope of the capacitance line is greatest. In contrast, when the long axis of the cantilever is perpendicular to the fast scan direction (a "90° scan," dashed line) the slope of the capacitance line is much smaller.

We also found that the slope of the linear variation in $C_{stray}$ is nearly independent of scan size and scan velocity. Figure 3.6(b) is an illustration of how the total change in capacitance $\Delta C$ depends on the scan size $\Delta x$. We see that $\Delta C$ increases linearly with scan size, suggesting that this variation is a geometrical effect and not caused by time-dependent effects such as piezoelectric "creep."
Figure 3.6: (a) 30μm long line scans on flat, uniform 10 nm thick SiO₂ film grown on Si. with fast scan direction parallel to (0° (solid line)) and perpendicular to (90° (dashed line)) the cantilever long axis. (b) Graph of capacitance change versus scan length for 0° (solid line) and 90° (dashed line) scans.

We attribute this variation in \( C_{stray} \) with lateral position to changes in the "tilt" of the probe assembly as it is scanned across the surface. This effect is illustrated in figure 3.7. First, consider the 0° scan case. The cantilever is moved laterally by a bending of the piezoelectric scanner tube. When the probe assembly is displaced towards the probe-tip end of the cantilever, the cantilever assembly becomes more parallel to the sample. This reduces the average separation between the probe assembly and the sample, hence increasing the stray capacitance (figure 3.7(a)). Conversely, when the probe assembly is displaced away from the probe-tip, the cantilever assembly becomes less parallel to the sample and a resulting increase in average separation and decrease in stray capacitance occurs (figure 3.7(b)). We have confirmed that the capacitance does increase when the probe is extended toward the probe-tip end of the cantilever.

Next, consider the 90° scan case where the cantilever is not precisely parallel to the surface (figure 3.7(c) and (d)). As the probe is scanned to the right, the left half
of the probe gets closer to the sample while the right half retracts. The variation in average separation is much smaller than the variation during a 0° scan and hence, the capacitance variation is much smaller. In fact, the observed variation in capacitance during 90° scans may well be caused by a slight misalignment of the long axis of the cantilever to the fast scan direction. This would add a contribution from the 0° scan component to the capacitance variation.

**Capacitance change with tip and holder height above sample**

The other change in stray capacitance that was observed during scanning was a change in capacitance as the cantilever–holder height above a sample was varied. We examined this effect by recording capacitance as force calibration curves were acquired. In these experiments, as shown in figure 3.8(a), the probe is positioned well
above the surface and the piezo is extended so that the tip approaches the surface. At 
a critical distance, attractive forces between the tip and surface cause the cantilever 
to abruptly deflect towards the surface and the tip makes contact with the surface 
(figure 3.8(b)). As the piezo continues to extend, the tip–surface interaction changes 
from attractive to repulsive and the cantilever now deflects upward (figure 3.8(c)). Next, the piezo reverses direction and retracts the cantilever from the surface. During 
retraction, adhesive forces keep the tip in contact with the surface and the cantilever 
bends downward for some time. Next, the tip snaps free of the surface, the cantilever 
returns to its undeflected position, and the retracting piezo continues to increase the 
tip height. An average of 25 deflection and capacitance versus piezo position curves, 
during both extension and retraction, are presented in figure 3.9. The sample was a 
10 nm thick thermal SiO$_2$ film grown on heavily doped ($10^{20}$/cm$^3$) n-type Si.

As the probe assembly approaches the surface (figure 3.9(a)), we see that the 
measured capacitance increases in a nearly linear fashion. However, as the tip nears 
the surface, the slope of the capacitance curve, $dC/dz$, increases slightly. If we use 
a parallel plate capacitor as a crude model of the probe assembly–sample geometry, 
we expect $C \sim A/z$, where $A$ is the effective plate area and $z$ is the effective plate 
separation. In this case, the slope of the capacitance–height curve should vary as 
$A/z^2$. Hence, we expect the slope $dC/dz$ to increase slightly as the tip approaches 
the surface. As approach continues, at the point the tip jumps into contact with 
the surface, there is an abrupt jump in capacitance. An expanded scale view of the 
snap-on region of the capacitance–height curve is shown in figure 3.9(b). Note that 
we can clearly see the $\sim 40$ aF change in capacitance caused when the tip jumps 
into contact with the surface. The capacitance jump does not appear abrupt because
Figure 3.8: Schematic of cantilever chip, cantilever, and probe tip during a force calibration curve. (a) Cantilever is undeflected cantilever with the tip out of contact during approach. (b) Cantilever beam deflects downward as attractive/adhesive forces pull the probe-tip onto the surface. (c) Cantilever beam deflects upward as repulsive forces between the tip and sample are caused by further extension of the z-piezoelectric tube.

the measurement time constant of the lock-in is longer than the sampling period. Finally, as the tip enters repulsive contact, the slope of the capacitance-height curve decreases. This is because the height of the tip end of the cantilever is now fixed, so the average height of the cantilever and holder decreases more slowly as the z-piezo is extended.

During retraction the capacitance changes are reversed. As the piezo pulls the cantilever, adhesive forces maintain contact with the surface, the cantilever deflects, and the capacitance decreases in a nearly linear fashion. When the force due the
Figure 3.9: (a) Force calibration curves and corresponding variations in $\Delta C$ as the tip approaches and retracts from the surface are shown. The slope of $\Delta C$ versus $z$ curve decreases when the tip is in contact. This is due to the separation of the cantilever from the surface being fixed at the probe-tip end (see text). Also note the sudden increase (decrease) in $\Delta C$ that occurs when the tip jumps to (snaps out of) contact. (b) An enlarged view of the increase in $\Delta C$ that occurs when the tip jumps to contact is shown.

bending cantilever exceeds the adhesive force, the tip snaps free of the surface and the capacitance decreases suddenly. This is caused by the sudden change in average cantilever height when the cantilever returns to its undeflected position. Finally, as the cantilever retracts away from the surface, the capacitance decreases in a nearly linear fashion.

The typical value measured for the slope of the capacitance–height curve, with the tip out of contact with the surface, is about 2 aF/nm for the MESP probes and 3.5 aF/nm for the CSC11/TiN probes. The CSC11/TiN cantilevers should have a larger $dC/dz$ than the MESP cantilevers because they are composed of two beams that form
a triangle while the MESP cantilevers are composed of a single beam. This results in a larger effective area $A$ for the CSC11/TiN cantilevers. The standard deviation of the measured slopes is approximately 10%. Likely causes of this variation include varying tip-sample interactions caused by adsorbed contaminants such as water, changing piezoelectric scanner tube response to applied voltage, and changes in the geometry of the probe assembly.

**Capacitance changes associated with topography**

We can now understand the anomalous capacitance data measured while scanning across a trench in Si and presented in figure 3.5. The observed slope in capacitance versus lateral position is mostly due to the probe–assembly tilt effect described in section 3.2.2. The anomalous, large increase in capacitance as the tip moves down into the trench requires more thought. First, an upper limit estimate of the tip-sample capacitance is obtained by approximating the tip-sample capacitance as a parallel plate capacitor with an area given by $\pi r_{\text{tip}}^2$, where $r_{\text{tip}}$ is the tip radius. For $r_{\text{tip}}$ of ~ 50 nm, the change in capacitance caused by the 100 nm change in SiO$_2$ film thickness encountered while scanning across the trench would only be ~ 25 aF. In contrast, the measured change is ~ 350 aF, or 15 times too large. Clearly, the film thickness variations cannot be responsible for the measured capacitance. Next, consider the change in probe assembly–sample geometry as the tip is scanned across the trench. In “constant force mode” AFM, the entire probe assembly is extended as the tip is scanned into the trench. During this extension, the cantilever deflection is kept nearly constant. This means that the average separation $h$ between the sample and, relatively large, probe assembly decreases by roughly the size of the step. This effect is shown schematically in figure 3.10 where a cross-section of a tip on a cantilever
is shown during a scan down into a trench. This decrease in average separation \( h \) causes the increase in capacitance shown in figure 3.5.

In figure 3.11, we show that the height sensitivity of the capacitance can be used to account for the variation in capacitance measured when scanning across the trench. This data was obtained with a 90° scan angle on the same trenched sample as the 0° scan data shown in figure 3.5. Line (a) in figure 3.11 is the topography trace after “flattening” the image by subtracting a linear least squares fit to the upper terraces of the trench. The dots (b) in figure 3.11 show the corresponding capacitance trace. It too has been flattened to remove the slope introduced by the lateral position dependent change in capacitance discussed in section 3.2.2. Finally, we estimate the change in capacitance \( \Delta C(x) \) caused by the change in the average probe-assembly height as 

\[
\Delta C(x) = (dC/dz)_{0}z(x).
\]

Here, \( (dC/dz)_{0} \approx -3.6 \text{ aF/nm} \) is the measured slope of the capacitance-height curve, evaluated where the tip is just out of contact, and \( z(x) \) is the measured topography. This estimate of \( \Delta C(x) \) is shown as the dashed line.
(c) in figure 3.11. It has essentially the same shape and magnitude as the measured \( \Delta C(x) \) but, in this case, over-estimates the measured \( \Delta C'(x) \) by about 10%. This is most likely due to an approximately 10% error in the determination of \( (dC'/dz)_0 \). We have repeated this procedure a number of times, across deep (\( \sim 100 \text{ nm} \)) trenches on multiple samples and in general find that the calculated \( \Delta C(x) \) agrees with the measured \( \Delta C(x) \) within \( \sim 10\% \).

The original aim of our research was to evaluate material properties, such as thin dielectric film thickness, with scanning capacitance microscopy. In the process of making SCM measurements, we discovered that scanning induced variations in parasitic capacitance are large but reproducible and mostly predictable. We find that about 90% of the topography induced variations in parasitic capacitance may be accounted for when making SCM measurements, as discussed in the previous section. Hence, we can estimate the range of sample topographical variations that would produce unaccounted topographical capacitance variations that are below the \( \sim 5 \text{ aF} \) noise level of our system. We estimate the unaccounted topographical capacitance variations as 10% of the calculated topographical variation. We then assume a maximum unaccounted capacitance variation of 5 aF and solve for \( \delta z \) as:

\[
\delta z = 5\text{aF}/(0.1 \left( \frac{dC}{dz} \right)_0).
\]  

With typical \( (dC/dz)_0 \) values between 2-4 aF/nm, topographical variations of less than 12 to 25 nm would produce unaccounted capacitance variations smaller than the 5 aF noise level. Hence, on sample surfaces with roughness of less than 12-24 nm, we can measure tip-sample capacitance changes down to the 0.35aF/\(\sqrt{\text{Hz}}\) noise level of our system after accounting for parasitic capacitances.

100
Figure 3.11: Line traces of (a) surface height and (b) $\Delta C$ during a $90^\circ$ scan across a 104 nm deep trench etched through SiO$_2$ on Si and covered with 10 nm Si$_3$N$_4$. The height scan was flattened to remove an average slope. The $\Delta C$ scan was flattened to remove the lateral position dependent slope in $C_{\text{stray}}$ (see text). Dashed line (c) was calculated by multiplying surface height line (a) by -3.6 aF/nm, the slope of the measured $\Delta C$ versus $z$ data.
3.3 Direct, low-frequency SCM measurements of dielectric film thickness

We have developed a circuit for scanning capacitance microscopy, characterized its performance, and quantified systematic variations in the parasitic capacitance. With this instrumentation and information at our disposal, we can address our original problem: how can SCM measurements be used to quantify thickness variations in thin, dielectric films? With our circuit, nm scale changes in capacitance between the tip and sample as small as 5 aF can be readily quantified as the tip is scanned over thin, dielectric films on conducting substrates. However, these capacitance variations cannot be directly related to film thickness variations until several steps are taken. First, the magnitude of just the nm-scale tip-sample capacitance must be extracted from the overall capacitance. Second, the relationship between the nm-scale capacitance and thin film thickness depends on the tip-sample geometry. This dependence is not simple and so it must be determined. Third, when scanning in air, a meniscus—which alters the tip-sample capacitance—is usually present: its effect on capacitance must be accounted for. We have used a combination of measurements and numerical simulations to address these issues and develop a means of calculating thickness variations from capacitance variations. There is a caveat: the meniscus must dominate the nm-scale tip-sample capacitance. That the meniscus dominates is not directly demonstrable from any single measurement presented here. However, when taken together, the variety of measurements presented lead us to conclude that the meniscus does dominate the nm-scale tip-sample capacitance in the experiments presented here.
3.3.1 Nanometer scale tip–sample capacitance measurement

The first challenge is to extract the nm-scale tip–sample capacitance from the total measured capacitance variation. In fact, the capacitance is a continuum and a working division of this continuum into nm-scale and long-range capacitance is necessary. The important portion of the capacitance is that portion which is responsible for probing the dielectric film: the portion contributed by the fields under the tip and within the dielectric. With these concepts in mind, we have chosen the capacitance change that occurs as the AFM tip snaps into contact with the surface as a rough definition of the nm-scale capacitance. That this is a reasonable definition is demonstrated in figure 3.12(b). First, there is a significant increase in capacitance when the tip snaps into contact. Second, the capacitance change prior to snap-on is very nearly linear. This indicates that it is dominated by: the portions of the tip away from the surface, the cantilever, and the probe assembly.

With this working definition of nm-scale capacitance, a method was developed to extract the tip-sample capacitance from simultaneously measured cantilever deflection and probe-assembly-sample capacitance as a function of probe-assembly height ("force-calibration curves"). An example force calibration curve, which is an average of 10 scans, is presented in figure 3.12. The sudden negative cantilever deflection pointed to by arrow (i) in figure 3.12(a) is the result of electrostatic and meniscus formation forces causing the cantilever to deflect and the tip to snap onto the surface. The cantilever deflection passes through 0—at the Δz value shown by the vertical dashed line—as the probe-assembly is extended further towards the surface.

The simultaneously acquired capacitance data is presented in figure 3.12(b). As the tip-cantilever-probe assembly approaches the surface the long range capacitance
Figure 3.12: Force calibration data is used to measure the nm-scale tip-sample capacitance. (a) is a plot of the cantilever deflection measured during approach. Note the sudden snap into contact (i) and the 0 deflection point while the tip is in contact with the surface (vertical dashed line). (b) is a graph of the simultaneously measured capacitance change which occurs during extension. The extrapolated long-range capacitance change caused by the cantilever and probe assembly (ii) is subtracted from the capacitance data to obtain (c), the graph of residual capacitance during extension. Note the sudden increase of capacitance that occurs at snap-on. $C_{ts}$ is the tip-sample capacitance which is measured when the cantilever deflection is 0.
increases monotonically. This is mostly due to the long range capacitance of the AFM probe assembly (cantilever, holder, wires) as it approaches the surface (section 3.2). The contribution of this long range capacitance is removed from the data by subtracting a quadratic polynomial fit to the out of contact, long range capacitance (figure 3.12(b)(ii)). The residual capacitance (figure 3.12(c)) consists of the tip-sample capacitance, including the meniscus contribution, and a changing capacitance caused by a changing cantilever deflection. An abrupt increase in the residual capacitance is observed when the tip snaps onto the surface (figure 3.12(iii)). This abrupt increase is caused by increased tip-sample interactions—including meniscus formation—and the change in cantilever deflection. Note however, that the change in residual capacitance caused by the change in cantilever deflection should be 0 where the cantilever deflection is 0. Now, when the tip is on the surface with 0 cantilever deflection, as marked by the vertical dashed line in figure 3.12, the residual capacitance is the tip-sample capacitance $C'_{ts}$. Hence, the nm-scale tip-sample capacitance—including the contribution of the meniscus—may be quantified when the tip is contacting a thin dielectric film.

### 3.3.2 Spherical tip model of tip-sample capacitance

We have measured the nm-scale tip-sample capacitance. Now, we must determine the relationship between capacitance and film thickness variations. This relationship depends on the geometry of the tip-sample system. The geometry of an AFM tip is complex but, it is often described in data sheets as a spherical tip at the end of either a pyramid or cone [76, 77]. For example, the MESP tips from Digital Instruments, are described as a $\sim 35$ nm radius tip on a 15° 1/2-angle cone that is 10µm
in length [76]. In reality the tip is not spherical, may have defects, and likely changes with scanning. Hence, in contrast to a parallel plate capacitor, analytic relationships between tip–geometry, film thickness, and capacitance are not available. Hence, numerical models of the tip-sample capacitance were used to estimate the magnitudes of the tip-sample capacitance for various tip sizes, dielectric film thicknesses, and tip-sample separations.

An AFM tip is difficult to model directly because there are two very different length scales (10 μm and ~1 nm). Numerical models rely on a grid of discrete points for calculations of the continuous electric fields. A grid appropriate for the long length scale would not be accurate for the short length scale while a grid appropriate for the short length scale would require too much memory to be implemented. When a subset of a conical tip model is used—by truncating the tip a few hundred nm above the surface, for example—another difficulty arises. The capacitance increases monotonically as the length of the cone increases [78]. This causes the same difficulty in extracting the tip-sample capacitance from either the model or the experimental data: the small tip-sample capacitance must be extracted from a large background capacitance. For these reasons, we choose to model the AFM tip-sample system as a sphere over a planar, thin film on a conducting substrate. We used a finite element partial differential equations solver, FlexPDE, to find a numerical solution to Laplace's equation. A diagram of the geometry and boundary conditions of the spherical model is presented in figure 3.13. The boundary condition at the system edges (shown as (i,iii,iv)) and the film-air interface (ii) is the continuity of the electric displacement \( \vec{D} \). Constant potential boundary conditions appropriate to conductors are applied to the sphere and the bottom of the film. The sphere is held at 1 V and
Figure 3.13: (a) A schematic of the system for a spherical capacitor model. Boundary conditions at: (i-iv) are continuity of \( \hat{D} \); tip surface, \( V=1 \), and bottom of film, \( V=0 \). (b) The calculated isopotential lines for a 100 nm sphere on a 50 nm film are shown.

the bottom of the film is held at ground. The small (0.05 nm) gap between the sphere and the film prevents an inconsistent grid from being generated.

We calculated the capacitance of this system for a variety of film thicknesses and sphere radii. The results of these calculations and a comparison with a parallel plate capacitor are presented in figure 3.14. This figure illustrates that the capacitance sensitivity to film thickness decreases with tip radius. Further, the comparison with
the parallel plate capacitance versus film thickness shows that spherical capacitors are much less sensitive to film thickness than are parallel plate capacitors. For this comparison, the area of the parallel plate capacitor was chosen to match the capacitance of the 300 nm radius tip at the 2.5 nm film thickness. There is one other take home message illustrated in figure 3.14. The sensitivity to film thickness of spherical capacitors disappears rapidly as film thickness increases. This suggests that SCM is most useful for measuring thickness variations in very thin films and that large radius tips are required for making these measurements. As an explicit example, the capacitance change when scanning a 300 nm radius sphere from a 2.5 nm to a 10 nm thick SiO\textsubscript{2} film on a conducting substrate is 14.8 aF. In comparison, the capacitance change when scanning from a 2.5 nm to a 10 nm thick SiO\textsubscript{2} film on a conducting substrate with a parallel plate capacitor is 92.9 aF. In this comparison the parallel plate capacitance over the 2.5 nm film is the capacitance of the sphere on the 2.5 nm film. It corresponds to a circular parallel plate capacitor of radius 53 nm.

3.3.3 SCM imaging of dielectric film thickness variations

In addition to numerical modelling of capacitance changes that should occur when scanning across variations in oxide thickness, we made measurements of capacitance changes that do occur when scanning across known variations in oxide thickness. A schematic of the sample is shown in figure 3.15(a). It was prepared by etching 3\textmu m wide trenches through a 10 nm thick thermal oxide and then allowing a 2.5 nm thick native oxide to form on the heavily doped \(5 \times 10^{17} - 10^{19} \text{cm}^{-3}\) n-type Si. The average thicknesses of both the 10 nm thermal oxide and the 2.5 nm native oxide were measured on large, uniform-thickness areas with ellipsometry.
Figure 3.14: Capacitance versus dielectric film thickness is plotted for 35, 50, 100, 200, and 300 nm radius spheres. The green line is the calculated capacitance of a parallel plate capacitor of 53 nm radius.

A contact mode AFM topography image and the corresponding corrected capacitance image obtained on this sample are shown in figure 3.15(b) and (c). Average cross-sections (128 lines) of these images are shown in figure 3.16. The nominal radius of the tip used to obtain these images is 100–300 nm [77]. The variation in measured capacitance associated with the variation in film thickness is easily observed, measured at high signal to noise, and quantified in capacitance units. As expected, the capacitance increases as the tip is scanned from the 10 nm thick thermal oxide to the thin, native oxide; surprisingly, the capacitance change is \( \sim 200 \text{ aF} \). In addition, we observe a fairly sharp decrease in capacitance when the tip is being scanned near the edge of a plateau. This may be due to enhanced electric fields at step edges [79], an interaction of the meniscus with the edge (see below), or both. In any case, the width of these features suggest a limit on the lateral resolution of this technique of \( \sim 200 \text{ nm} \).
Figure 3.15: (a) A schematic cross section of the SiO$_2$/Si trench sample is shown. (b) An AFM topography image of the trench is shown. The full grey-scale range is 10 nm. (c) The corrected (see text) capacitance image of the trench is shown. The full scale is $\sim 200$ aF.
The qualitative behaviour of the capacitance while scanning the tip across the sample makes sense. In order to quantify the capacitance and relate it to dielectric film thickness we must first account for the linear change in capacitance caused by the tilting of the cantilever as the tip is scanned and the change in average cantilever height induced by scanning across topography features. These changes in parasitic capacitance are detailed in section 3.2. As is usual in AFM imaging, the topography image was flattened by subtracting a best fit quadratic polynomial to the plateaus of each scan line from each scan line. The capacitance image correction procedure is illustrated in figure 3.16(b) and is summarized as follows. First, a best fit line to the plateaus of each scan line was subtracted from each scan line to remove an average slope. The average slope in parasitic capacitance is caused by the change in the tilt of the cantilever assembly as the probe is scanned. Second, the calculated changes in parasitic capacitance caused by the changes in average tip height—shown as line (ii) in figure 3.16(b)—as the probe is scanned across topographical features is subtracted from the measured capacitance. Note that this correction is a small fraction of the measured capacitance change. The resulting corrected capacitance is shown in line(iii) of figure 3.16(b) and in the capacitance image (figure 3.15(c)).

There is a surprising aspect to the measured capacitance data shown in figure 3.15 figure 3.16. The change in capacitance (~ 200 aF) observed when scanning across the 2.5 nm thick oxide region is anomalously large when compared to the capacitance change predicted for a 100–300 nm radius tip. As noted in section 3.3.2, a spherical capacitor model of an AFM tip on a thin oxide film predicts a change of only 4.5 aF for a 100 nm radius tip and only 14.8 aF for a 300 nm radius tip.
Figure 3.16: (a) is the average cross-section of all topography lines in figure 3.15(a), line (i) in (b) is the average of all raw capacitance data lines used to produce figure 3.15(b). The dashed line (ii) is the calculated topography induced parasitic capacitance change while the dotted line (iii) is the corrected capacitance displayed in figure 3.15(b).
We suggest that increased sensitivity to film thickness occurs because of the presence of an aqueous meniscus. The high dielectric constant of water (\( \sim 80\epsilon_0 \)) compared to both air (\( \sim \epsilon_0 \)) and SiO\(_2\) (3.9\( \epsilon_0 \)) causes the electric field to be confined to the surface of the meniscus and increases the coupling between the tip and sample. In essence, the tip-sample capacitance behaves more like a parallel plate than a spherical capacitor [80]. This enhances the sensitivity of the tip-sample capacitance to dielectric film thickness.

Along with the enhanced sensitivity to film thickness comes a complication: the meniscus area must be known and verified to be constant during the measurements. In general, the effective area \( A \) is not known and the problem becomes how does one determine \( A \)? For a parallel plate capacitor, the effective meniscus area \( A \) may be calculated from

\[
A = \frac{C_{ts} t_o}{\epsilon}
\]

if the tip-sample capacitance \( C_{ts} \) (including the effect of the meniscus), the dielectric thickness \( t_o \), and the dielectric constant of the film \( \epsilon \) are all known.

3.3.4 Environmental effects on meniscus size

Several questions arose when we first considered the hypothesis that a meniscus increases the sensitivity of capacitance to dielectric film thickness. (1) How do we know that there is a meniscus present? (2) Is the effective, electrical area of a meniscus correlated with the adhesive force caused by the meniscus? (3) Is the enhanced sensitivity a function of humidity and other environmental factors? (4) Does the
meniscus size change while scanning and if it does, can we detect the change as being distinct from film thickness changes?

The first question is easiest to answer. We know there is a meniscus present by examining the force calibration curves. A meniscus causes a very large adhesive force which is observed as a much larger snap-off deflection than snap-on deflection. Since cantilever deflection is nearly linearly related to the force on the cantilever, the snap-off force is much larger than the snap-on force. We always observe differences, often quite large, in snap-on and snap-off deflection. This difference between snap-on and snap-off deflections is illustrated in figure 3.17 where the snap-off deflection is $\sim 5$ times the snap-on force. This is not sufficient because any distance dependent force would exhibit this characteristic. However, the size of the snap-off deflection varies, as will be discussed later, with environmental influences including time, humidity, and sample history.

The second question—is the effective, electrical area of the meniscus correlated with the meniscus adhesive force—was addressed by plotting the measured, nm-scale tip-sample capacitance (section 3.3.1) against the snap-off deflection. The force calibration curve acquired at the same time as the tip-sample capacitance measurement was used to determine the snap-off deflection. Figure 3.18 shows that the two are in fact correlated. Note that the snap-off deflection was not a controlled parameter. However, we could make general meniscus size (snap-off deflection size) selections by controlling ambient humidity and sample preparation.

That we can make general meniscus size selections by controlling the environment addresses the third question: is the enhanced sensitivity to meniscus formation a function of environment. We know that meniscus formation and behaviour depends
Figure 3.17: The cantilever deflection as a function of piezo-electric scanner tube displacement in $z$ is shown here for extension (blue line) and retraction (red line). Note the different snap-on and snap-off deflections and hence, forces.

Figure 3.18: The circles represent individual measurements of tip-sample capacitance $C_{ts}$ plotted against the snap-off cantilever deflection and the dashed line is the linear regression (constrained to a 0 intercept) for that data. The correlation is indicative of a relationship between meniscus forces and meniscus effective electrical area.
upon environmental factors such as relative humidity. surface chemistry, and geometry so we performed a series of experiments to explore the influence of environment on tip-sample capacitance [81]. We measured the tip-sample capacitance from force calibration curves at a variety of different humidities and for several different sample preparation methods. Figure 3.19 is a graphical summary of this work. The principle conclusion is that the size of the meniscus—as evaluated with snap-on capacitance and snap-off cantilever deflection—does not depend solely on humidity. However, the size of the meniscus does depend on the sample history coupled with humidity. We observed that the smallest menisci were formed immediately after baking the sample at 150°C for 30 min. The measured snap-off deflection increased in time slowly if the ambient environment was at low (< 5% RH) relative humidity. In contrast, the snap-off deflection quickly stabilized at a large value when the ambient humidity was higher (~ 40% RH). Further, if a “baked” sample was exposed to higher relative humidity, the meniscus size did not decrease when the humidity was reduced to < 5% RH. However, small menisci could be recovered by re-baking the sample.

The last question about menisci in capacitance measurements that we address is: does the meniscus size change while scanning and if it does, can we detect the change as being distinct from film thickness changes. In order for us to exploit the enhanced sensitivity caused by the presence of a meniscus, the effective area of the meniscus must be constant during a scan. We have two observations that address this issue. First, we measured the tip-sample capacitance before and after obtaining SCM images. In most cases, the tip-sample capacitance did not change more than 5–10%. Second, when the tip-sample capacitance did change significantly, we observed an abrupt change (between adjacent lines) in sensitivity during imaging.
Figure 3.19: This is a summary of the affect of environment on tip-sample capacitance. The squares correspond to capacitance measured from the snap-off data and the circles to capacitance measured from the snap-on data.
The tip-sample capacitance measured before the sensitivity change and the thickness calculated from data obtained before the change were consistent. Similarly, the tip-sample capacitance measured after the sensitivity change and the thickness calculated from data obtained after the change were consistent. Based on these observations we conclude that the meniscus is reasonably constant during imaging and that changes in the meniscus that occur during imaging are distinguishable from film thickness changes.

3.3.5 Calculating dielectric film thickness from SCM measurements

Variations in film thickness may now be calculated from variations in capacitance along a scan line $\delta C_{ts}(x)$ from

$$ t(x) = \frac{t_o}{1 + \frac{\delta C_{ts}(x)}{C_{ts}}} $$(3.6)

where $t_o$ is the average oxide thickness on the plateau and $C_{ts}$ is the measured tip-sample capacitance. An example of this calculation is presented in figure 3.20. The red line is the SiO$_2$ film thickness calculated from equation 3.6 and the assumption that $t_o$ is the average film thickness (10 nm) on the plateau. The blue line is a schematic of the film thickness profile based on ellipsometry measurements of the 10 nm and 2.5 nm thick films. We see that the calculated thickness is within $\sim 1$ nm of the known average thickness.

This is a $\sim 40\%$ overestimate of the oxide thickness. There are two effects that could be contributing to the overestimate of oxide thickness. First, the sample substrate is Si, a semiconductor not a metal. This implies that there could be a voltage dependent depletion region near the SiO$_2$/Si interface and that the capacitance measured
includes this layer. The effect of such a layer would be greater when scanning over the 2.5 nm oxide than when scanning over the 10 nm oxide. This would cause the capacitance change when scanning from the thick to thin oxide to be less than one would predict from the oxide thickness alone. Second, the meniscus may not completely dominate the tip-sample capacitance. In this case, the capacitance from the tip would cause a smaller capacitance change when crossing from the thick to thin oxide than would be expected for a parallel plate capacitor. Hence, the calculated oxide thickness would be overestimated.

3.3.6  Procedure Summary and Discussion

The procedure developed for measuring the nm-scale tip-sample capacitance is as follows. First, force calibration curves are obtained and the capacitance variations
that occur when the tip is farther than $\sim 20$ nm above the surface are fit with a least squares quadratic polynomial. This polynomial is then subtracted from the capacitance data. The residual capacitance contains the nm-scale tip-sample capacitance and the capacitance change caused by the changing cantilever deflection. The capacitance change caused by the changing cantilever deflection is removed by evaluating the residual capacitance at the point where the cantilever deflection is zero and the tip is on the surface. At this point, the residual capacitance is the nm-scale tip-sample capacitance. The nm-scale tip-sample capacitance includes the effect of any meniscus. In fact, we assume that the meniscus dominates the tip-sample capacitance so that the tip-sample capacitance is very nearly that of a parallel plate capacitor. With this assumption and the known average thickness of the dielectric film we calculate the effective, electrical area of the meniscus. Finally, variations in film thickness are calculated from: capacitance variations measured while scanning across a sample with film thickness variations: the effective, electrical area of the meniscus: the known average film thickness for a region of the film: and a parallel plate capacitor model.

If this procedure is to be of general use, the question becomes: is our assumption that the meniscus dominates the tip-sample capacitance valid? If the tip by itself contributes substantially to the capacitance, our analysis of the capacitance variations when crossing a step is incorrect. In contrast, the more the meniscus dominates the capacitance, the more correct our analysis is. I will argue that the meniscus is responsible for the bulk of the measured tip-sample capacitance for the data shown in section 3.3.3 and that the depletion layer effects cause the overestimate of the 2.5 nm oxide film thickness.
Figure 3.21: The variation in capacitance for 100, 200, and 300 nm radius spheres as a function of height above a 10 nm SiO$_2$ film is presented. Note both the long and short range components of the capacitance.

There are three principle arguments that the meniscus dominates the tip-sample capacitance: first, the variation in capacitance of a sphere as a function of height suggests that the tip itself can contribute no more than $\sim 50\%$ for a 300 nm radius tip and $\sim 10\%$ for a 100 nm tip; second, the presence of a meniscus reduces the contribution from the tip by "shorting" the bulk of the energy density in the electric field; and third, that the depletion capacitance contribution can account for the bulk of the observed overestimate of the 2.5 nm oxide thickness.
First, consider the variation in capacitance of a sphere as a function of height for 100, 200, and 300 nm radius spheres above a 10 nm SiO$_2$ film (Figure 3.21). These curves show that there are two contributions to the capacitance: long-range, slowly decreasing and short-range, rapidly decreasing components. Consider an isolated sphere to understand the long range component. The voltage at the surface of the sphere $V$ is given by

$$V = \frac{Q}{4\pi\varepsilon_0 a} \quad (3.7)$$

where $Q$ is the charge on the sphere, $\varepsilon_0$ is the dielectric constant of a vacuum, and $a$ is the radius of the sphere. Now, the capacitance $C$ of an isolated sphere is derived from equation 3.7 and the definition of capacitance as $C \equiv Q/V$. It is

$$C = 4\pi\varepsilon_0 a \quad (3.8)$$

This equation says that the capacitance of a sphere as it is moved towards infinity will never decrease below that given by equation 3.8. Further, as seen in figure 3.21, the capacitance at $\sim 20$ nm above the surface is only 32 aF less than that of the capacitance on the surface. This means that approximately 70% of the contribution of the tip to the total capacitance it is included in the long range capacitance. Hence, only $\sim 30\%$ of the tip capacitance contributes to the tip-sample capacitance we have defined by the capacitance increase over "baseline" that we measure with a force calibration curve.

Note that this is not inconsistent with the idea that there is a nm-scale tip-sample capacitance. The long-range capacitance, even that from only the tip, does not "sample" the thin, dielectric film. To understand this, consider the following: the
connected to the idea that the long-range capacitance does not sample the thin dielectric film is my argument that the meniscus reduces the contribution from the tip. First, consider a small meniscus. As shown in figure 3.22, the a small meniscus makes electrical contact with the tip. Since the dielectric constant is so high, the outside surface of the meniscus is essentially an isopotential surface. This causes an isopotential region to form between the sphere and the film surface. Therefore, there is no voltage drop in the air between the meniscus and tip so there is only a small electric field in this small region. Consequentially, the region between the tip and meniscus no longer contributes significantly to the tip capacitance. A large meniscus also creates a large isopotential region that does not contribute to the tip capacitance.

Now, consider the electric energy density under a sphere in contact with the film when there is no meniscus present. Virtually all of the energy density is concentrated within 50 nm of the contact point as is shown in figure 3.23. This entire region is in the area converted into a nearly iso-potential region by the meniscus. Hence, as we assumed in our analysis, the meniscus controls the tip-sample capacitance.

Finally, the tip-sample capacitance measurements we have made include the effect of a variable depletion region in the Si substrate. The depletion region reduces the capacitance below that of the oxide capacitance. Further, the depletion region is varied more by the AC voltage applied to the tip when the tip is over a 2.5 nm native oxide than when the tip is over a 10 nm thermal oxide. Hence, the 2.5 nm oxide film
Figure 3.22: A schematic of small and large menisci shows the large, nearly isopotential regions around a sphere that are created by a meniscus. The regions where the electric fields are modified by the presence of a meniscus are shown in yellow: the meniscus is shown in light blue.

appears electrically thicker than it would over a metal substrate. Preliminary measurements of the effective capacitance over a 10 nm region coupled with calculations of the high frequency CV characteristics for 10 nm and 2.5 nm oxides over the Si suggest that this effect can account for all of the oxide thickness overestimate.

We conclude—for the SiO$_2$ on Si system—that: the meniscus does dominate the nm-scale tip-sample capacitance, we have measured the electrical area of the meniscus for the data shown in figure 3.12, and that we can use that measurement to calculate thickness variations of SiO$_2$ on Si films. In contrast, previous SCM attempts to measure dielectric film thickness were not quantitative and relied on $dC/dV$, which requires a semiconducting substrate [65, 67]. Further, we anticipate that for thin dielectric films on conducting rather than semiconducting substrates, the depletion
Figure 3.23: The electric field energy density for a 100 nm sphere on a 10 nm SiO₂ film is shown.

layer effects would not be present and the thickness measurements would be more accurate. In these systems, we anticipate that sub-nm vertical resolution is achievable. However, if there is not a meniscus present or the meniscus does not dominate the nm-scale tip-sample capacitance, then variations in capacitance observed while scanning cannot be easily related to variations in thickness.
CHAPTER 4

CONCLUSIONS AND ROADMAP TO FUTURE WORK

An investigation into thin, dielectric films using electrical characterization techniques was performed. Two particular systems were studied. The first system is the charge trapping defect, or interface trap, found at the interface between Si and thermally grown nm-thick silicon dioxide (SiO$_2$) films grown on Si. The goal of this work was to make both ballistic electron emission microscopy (BEEM) and charge pumping measurements on the same device. These types of nm-scale spatially resolved measurements cannot be made on commercial MOSFETs because BEEM measurements require thin (~5nm) very clean metal film gates. In contrast, commercial devices are produced with thick poly-crystalline Silicon (p-Si) gates. As a first step in producing a BEEM compatible MOSFET, we have shown that BEEM measurements can be made on metal-oxide-semiconductor (MOS) capacitors that have been subjected to standard semiconductor fabrication processes. Further, through extensive fabrication process development, we have identified the critical technologies required to produce working MOSFETs that are compatible with both BEEM and charge pumping measurements. Finally, we are engaged in a collaboration with IMEC in Leuven, Belgium to produce, test, and use working MOSFETs in BEEM and charge pumping measurements.
The principle results of the Si/SiO$_2$ interface work is the identification of critical processing steps in the fabrication of BEEM compatible MOSFETs. First and foremost, the gate oxide must be protected through as much of the process as possible. For example, the gate oxide should be buried under a p-Si layer (~100 nm) as soon as the gate oxide is grown to protect the gate oxide during all subsequent processing steps. Second, use the p-Si layer as the As implant mask so that the wafers will not have to be exposed to an O$_2$ plasma for resist ashing. This may be a minor benefit because the p-Si should protect the gate oxide surface from the O$_2$ plasma in any event. Third, following the contact metallization steps use LPCVD to deposit an oxide passivation layer. This layer can serve as a mask against a chemical etch for the p-Si removal in the gate region so that plasma etching does not have to be used to remove the p-Si.

The collaboration with IMEC involves two distinct fabrication processes. In both processes, passivation layers are used to allow for chemical instead of plasma etch removal of the p-Si layer covering the gates. One of the processes is a self-aligned gate process while the other is not. The critical layers in the mask set for the self-aligned gate process were designed by us in close collaboration with Dr. B. Kaczer of IMEC. The other mask set was designed at IMEC with input from us on device dimensions and layout. We have received some partially processed samples from IMEC that have our devices on them. While these are incomplete devices, they allow us to evaluate the processing steps that will be used to make them BEEM compatible.

One device on these wafers that is of particular interest to us is the 1/2-MOSFET that is pictured in figure 4.1. This is a MOSFET that is made without a drain, but with a very large area gate region. Charge pumping studies will be difficult.
Figure 4.1: This is a picture of the 1/2 MOSFET structure that our collaborators at IMEC have made. Notice the large, planar gate region that allows easy access for our grounding electrode and our STM tip. This is not a working device, but is provided to us for the purpose of testing our modification processes.

to interpret with this device because there will be a large geometric component to the charge pumping current (section 2.2.4). However, the range of possible BEEM measurements is expanded by such a device. For example, we can use BEEM to inject electrons into the device at a variety of distances from the source and measure the resulting BEEM current in the source and substrate independently. This should allow us to create a nm-scale map of hot carrier transport within a working device.

The second system under study is the use of capacitance measurements to resolve sub-nm variations in the thickness of thin dielectric films with nm-scale lateral resolution. Towards this goal, we have: developed direct, low-frequency scanning capacitance microscopy (SCM) instrumentation capable of measuring $10^{-18}$F changes in the capacitance between an atomic force microscope (AFM) tip and a sample with
a noise level of 0.4 aF/√Hz: for the first time, quantified and developed means of account­ing for changes in parasitic capacitance that occur while scanning an AFM tip; quantified the effective area of the meniscus that forms between the AFM tip and the sample while scanning in air: and measured nm-scale dielectric film thickness variations with up to ~ 1 nm resolution with lateral resolution of ~ 200 nm.

The next step in utilizing SCM for thin film measurements is to try and improve the quantification of the meniscus effective area. Two approaches are recommended. The first approach is to use a SiO₂ on Si grating sample with a thermal oxide being grown in place of the native oxide. This would allow macroscopic CV measurements to be used to quantify the effect the voltage dependent depletion region has on the SCM measurements. The second approach is preferred. Develop a grating type sample with two different thickness films on a conducting rather than semiconducting substrate. This would eliminate the uncertainty introduced by voltage dependent CV characteristics.

Further studies of the lateral variations in electrical properties of fully bonded perfluoropolyether films would also be interesting. Measurements made during this study indicate that there are patches of conducting material. It would be interesting to learn if this is due to structural variations in the film or adsorbed water layers making contact with the surface. My speculation is that there are patches of the film that have adsorbed water layers that contact the underlying conducting diamond like carbon film on Al substrate because the contact forces used in these measurements are low. Making reliable contact to Au films with a conducting tip requires large contact forces.
APPENDIX A

FABRICATION RECIPES FROM THE OSU CLEANROOM

This appendix contains a series of recipes for the cleaning, lithography, oxidation, wet etching, and metallization steps. There are a couple of warnings to be noted before using any of the cleaning recipes. In more modern processing, both the initial degrease and the SC1 solution have been replaced with piranha etches. These are based on the self-heating solutions of $\text{H}_2\text{SO}_4:\text{H}_2\text{O}_2$ which are tabulated in Appendix B. Further, lower concentrations of HF for longer times are now commonly used for cleaning steps because higher HF concentrations result in markedly rougher surfaces. The rougher surfaces have higher densities of electrical defects \cite{83}.

One more warning about the cleanroom facilities in electrical engineering. These facilities are shared by many users who are performing fabrication tasks with a wide variety of materials. Many of these materials are incompatible with Si processing so the likelihood of contamination is high. In particular, the rinse sink in the “Gold” room is known to be contaminated with the following species: As, Au, Cu, Fe, and Ga. At various stages in Si processing exposure to any of these species can markedly diminish device performance. I would recommend that you use your own dedicated rinse containers for the deionized water (DI) rinses.
### A.1 Cleaning Recipes

Wafer cleaning is one of the most critical steps in semiconductor fabrication. The initial wafer clean consists of an initial degrease (table A.1) followed by an oxidation clean (table A.2). Cleaning is also important following photolithography steps: the recipe is given in table A.3.

### A.2 Oxidation Recipes

The following tables contain the recipes used for oxidation procedures in the OSU cleanroom. The diffusion oxidations were performed in the top tube, the field oxidations were performed in the middle tube, and the gate oxidations were performed in the bottom tube. It should be noted that the bottom tube is reserved for gate oxidations. “Cleaning” of the bottom tube was accomplished by bubbling O₂ through concentrated HCl for ~ 30 min prior to the oxidation. This process helps remove Na and Fe from the tube surfaces.
<table>
<thead>
<tr>
<th>Chemicals</th>
<th>time</th>
<th>temperature (°C)</th>
<th>notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>DI:HF 1:1</td>
<td>15 s</td>
<td>RT</td>
<td>*</td>
</tr>
<tr>
<td>DI</td>
<td>5 m</td>
<td>RT</td>
<td>rinse step</td>
</tr>
<tr>
<td>DI:NH₄OH:H₂O₂ 5:1:1</td>
<td>10 m</td>
<td>70 ± 5</td>
<td>SCI</td>
</tr>
<tr>
<td>DI</td>
<td>5 m</td>
<td>RT</td>
<td>rinse step</td>
</tr>
<tr>
<td>DI:HCl:H₂O₂ 5:1:1</td>
<td>10 m</td>
<td>70 ± 5</td>
<td>SC2</td>
</tr>
<tr>
<td>DI:HF 1:1</td>
<td>15 s</td>
<td>RT</td>
<td>*</td>
</tr>
<tr>
<td>DI</td>
<td>5 m</td>
<td>RT</td>
<td>rinse step</td>
</tr>
<tr>
<td>N₂</td>
<td>-</td>
<td>RT</td>
<td>blow dry</td>
</tr>
</tbody>
</table>

Table A.2: OSU preoxidation cleaning recipes. This recipe is a variant of the industry standard “RCA clean”. SCI is standard clean 1, SC2 is standard clean 2, and RT is room temperature. The HF dips marked with * are not used on gate oxides.

### A.3 Photolithography

Photolithography is used to protect some areas of a wafer from an etch or a material deposition while allowing other areas to be exposed to the treatment. Details of photolithography chemistry, physics, and most importantly, terminology, can be found in *Process Technology*. Vol. I of *Semiconductor Processing for the VLSI Era* [54]. A few general tips that are not often found in the textbooks follow. First, for best results the wafers should be clean. Second, it is best to apply resist within 30 min of a high temperature furnace step to assure that there is not an adhered water layer. Alternatively, a 30 min bake at 150°C will remove an adhered water layer [57]. Third, a common practice in the OSU cleanroom is to strip resist by spraying acetone on a spinning wafer and then blowing the wafer dry with photoresist. This procedure is fine if your minimum feature size is many microns. For smaller features or dense
<table>
<thead>
<tr>
<th>Chemicals</th>
<th>time</th>
<th>temperature (°C)</th>
<th>notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>acetone</td>
<td>1 m</td>
<td>RT</td>
<td>in ultrasonic bath</td>
</tr>
<tr>
<td>fresh acetone</td>
<td>5 m</td>
<td>RT</td>
<td>in ultrasonic bath</td>
</tr>
<tr>
<td>methanol</td>
<td>5 m</td>
<td>RT</td>
<td>in ultrasonic bath</td>
</tr>
<tr>
<td>DI</td>
<td>5 m</td>
<td>RT</td>
<td>rinse step</td>
</tr>
<tr>
<td>N₂</td>
<td></td>
<td></td>
<td>blow dry</td>
</tr>
</tbody>
</table>

Table A.3: OSU photoresist strip and clean. (RT is room temperature.) Note that the ultrasonic baths are not used when there is a thin, metal film on the wafers.

device packing, the wafer should be stripped of photoresist, cleaned, and baked before reapplying resist.

The general procedure for applying resist that was used in the OSU cleanroom is as follows. Several drops of hexamethyl-disilazane (HMDS), which improves the bonding of resist to the wafer, are applied to the wafer via a pipette and the wafer is spun at 5000 rpm (10000 rpm/s acceleration) for 10s. Next, for most of the work discussed in this dissertation, a small pool Shipley-1818—a positive photoresist—was applied to the wafer with a filtered syringe and the wafer was spun at 5000 rpm (10000 rpm/s acceleration) for 30s. For lift-off lithography, the wafer was spun at 4000 rpm (10000 rpm/s acceleration) to get a slightly thicker resist layer.

Following resist application the wafers are soft-baked in an oven for 30 min at 110 ± 5°C. This time is longer and temperature is higher than is used in EE-637 to assure solvent removal from the resist. This is especially important for lift-off processes [84]. A 1 min bake on a hot plate at 115°C may be substituted for the oven bake. Care must be taken to assure that the temperature remains below 125°C or the photosensitivity will be reduced [84].
<table>
<thead>
<tr>
<th>Procedure</th>
<th>Temperature °C</th>
<th>Time (min)</th>
<th>Gases</th>
<th>Flow rate (sccm)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Field Oxidation (~ 0.6 μm)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>push</td>
<td>1100</td>
<td>1</td>
<td>N₂</td>
<td>4370</td>
</tr>
<tr>
<td>dry</td>
<td>1100</td>
<td>30</td>
<td>O₂</td>
<td>2000</td>
</tr>
<tr>
<td>wet</td>
<td>1100</td>
<td>50</td>
<td>O₂</td>
<td>2000</td>
</tr>
<tr>
<td>dry</td>
<td>1100</td>
<td>24</td>
<td>O₂</td>
<td>2000</td>
</tr>
<tr>
<td>anneal</td>
<td>1100</td>
<td>10</td>
<td>N₂</td>
<td>4370</td>
</tr>
<tr>
<td>pull</td>
<td>1100</td>
<td>1</td>
<td>N₂</td>
<td>4370</td>
</tr>
<tr>
<td><strong>Thin Gate Oxidation (~ 10 nm)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>push</td>
<td>900</td>
<td>1</td>
<td>N₂</td>
<td>4370</td>
</tr>
<tr>
<td>dry</td>
<td>900</td>
<td>20</td>
<td>O₂</td>
<td>2000</td>
</tr>
<tr>
<td>anneal</td>
<td>1100</td>
<td>10</td>
<td>N₂</td>
<td>4370</td>
</tr>
<tr>
<td>pull</td>
<td>1100</td>
<td>1</td>
<td>N₂</td>
<td>4370</td>
</tr>
<tr>
<td><strong>Thick Gate Oxidation (~ 0.125 μm)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>push</td>
<td>900</td>
<td>1</td>
<td>N₂</td>
<td>4370</td>
</tr>
<tr>
<td>dry</td>
<td>900</td>
<td>59</td>
<td>O₂</td>
<td>2000</td>
</tr>
<tr>
<td>anneal</td>
<td>1100</td>
<td>14</td>
<td>N₂</td>
<td>4370</td>
</tr>
<tr>
<td>pull</td>
<td>1100</td>
<td>1</td>
<td>N₂</td>
<td>4370</td>
</tr>
</tbody>
</table>

Table A.4: OSU oxidation recipes: wet oxidations were done by bubbling O₂ through boiling deionized water.

Resist exposure was done with the Cobilt Aligner. Exposure times must be checked regularly with this aligner because of variations in UV-light intensity that occur with time. I found that 10–15 s was a good starting point. For lift-off lithography, the wafers are soaked in chlorobenzene soak (15 min) after exposure and before developing. This hardens the surface of the resist against the developer and causes undercutting of the resist layer during developing (Figure 2.9). The wafer is blown dry with N₂ prior to developing because the chlorobenzene is immiscible in water and forms droplets in the developer if not removed. Developing is done in Microposit
MF-320 for 1 min at room temperature. Finally, the wafer is hard baked for 20 min at 120 ± 5°C.

Metal removal during a lift-off step is performed by first soaking the metallized resist for 15 min in N-methyl-2-pyrolidone to swell the resist and loosen the metal. The wafers are then soaked in acetone, sprayed with acetone, and rinsed in DI. The normal resist strip and clean procedure follows for thick metal films. However, for thin metal films, the ultrasonic bath will strip the metal from the surface so the wafers are soaked in the chemicals with gentle agitation.
A.4 Metallization

Several different metallization procedures were tried. Thermal evaporations were the initial choice of metallization technique because e-beam evaporation generates x-rays which can cause damage to gate oxides. The NRC thermal evaporator in the clean-room was used to deposit Al. Al wire (~ 25 cm) was cleaned in 1:1 HCl:DI for ~ 30 s, rinsed in DI, and loaded into a W boat. The wafer was dipped in 10:1 DI:HF for 10s to remove native oxide in the contact region, rinsed in DI, and immediately loaded into the vacuum chamber. Note that for lift-off metallization, the gate oxide is protected against the HF etch by photoresist. The chamber is pumped down for > 1 hr since the ion gauge does not work and the base pressure is not known. Deposition was performed by ramping the boat heating current to ~ 2 A until the Al melts. The current is then increased rapidly to ~ 8 A. The thickness of the deposited metal is monitored with a quartz mass balance.
APPENDIX B

FABRICATION RECIPES FROM THE STANFORD NANOFABRICATION FACILITY

B.1 Starting Materials

All processing at the SNF was done on p-type, prime grade, 4" wafers obtained from the SNF stockroom. The wafers were ~ 0.5 mm thick with 8-11Ω-cm resistivities.

B.2 Wafer Cleaning

The initial clean and the photoresist step clean is a piranha etch. The wafers are immersed in a self-heating solution of concentrated sulfuric acid and hydrogen peroxide, rinsed with DI in the dump rinser, and finally rinsed and dried in the spin rinser. A hydrofluoric acid (HF) dip followed by a DI rinse in the dump rinser may be added after the dump rinser step. Details of chemical concentrations are presented in Table B.1.

Prior to any furnace step, other than the forming gas anneal, all wafers must undergo a diffusion clean. This starts with a modified piranha clean (Table B.1 diffusion piranha) to remove organic contaminants followed by a DI rinse. Next is an
<table>
<thead>
<tr>
<th>Cleaning Step</th>
<th>Bath Composition</th>
<th>Temperature</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial Piranha clean</td>
<td>9:1 H₂SO₄:H₂O₂</td>
<td>120°C</td>
<td>20'</td>
</tr>
<tr>
<td>Dump Rinse</td>
<td>DI</td>
<td>20°C</td>
<td>6 cycles</td>
</tr>
<tr>
<td>Spin Rinse/Dry</td>
<td>DI, N₂</td>
<td>heated</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>280s rinse. 120 s dry</td>
</tr>
<tr>
<td>Final Piranha clean</td>
<td>4:1 H₂SO₄:H₂O₂</td>
<td>90°C</td>
<td>10'</td>
</tr>
<tr>
<td>Dump Rinse</td>
<td>DI</td>
<td>20°C</td>
<td>6 cycles</td>
</tr>
<tr>
<td>HF dip</td>
<td>50:1 DI:HF</td>
<td>20°C</td>
<td>30s *</td>
</tr>
<tr>
<td>Dump Rinse</td>
<td>DI</td>
<td>20°C</td>
<td>6 cycles</td>
</tr>
<tr>
<td>Metal Scavenge</td>
<td>5:1:1 DI:HCl:H₂O₂</td>
<td>70°C</td>
<td>10'</td>
</tr>
<tr>
<td>Dump Rinse</td>
<td>DI</td>
<td>20°C</td>
<td>6 cycles</td>
</tr>
<tr>
<td>HF dip</td>
<td>50:1 DI:HF</td>
<td>20°C</td>
<td>30s *</td>
</tr>
<tr>
<td>Spin Rinse/Dry</td>
<td>DI, N₂</td>
<td>heated</td>
<td>15MΩ-cm, 120s dry</td>
</tr>
</tbody>
</table>

Table B.1: Standard Cleaning Procedures. HF dips marked with * are not used on unprotected gate oxides.

optional HF dip to remove the native oxide grown by the piranha clean followed by a DI rinse. Metal ions are then removed with an SC2 clean (5:1:1 DI:HCl:H₂O₂) followed by a DI rinse. Final rinsing is performed in the spin rinser with the endpoint for the rinse being a 16MΩ-cm resistivity of the rinse water. The wafers are dried in the spin drier and must be used within one hour or the diffusion clean must be repeated prior to introducing the wafers into any clean equipment. Details of these cleaning steps are presented in Table B.1.

### B.3 TYLAN furnace recipes

The complete TYLAN furnace recipes used for oxidation, nitride deposition, p-Si deposition, and anneals are currently (1 May 2002) available on the stanford web site under Equipment:Tylan Recipes [57]. They are essentially computer programs detailing each action performed by the furnaces. Where each process is used in the
<table>
<thead>
<tr>
<th>Step</th>
<th>Parameter</th>
<th>Push</th>
<th>Temperature Ramp</th>
<th>Growth</th>
<th>Cool</th>
<th>Pull</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pad Oxidation</td>
<td>Time Temp. Gases (sccm)</td>
<td>20' 800°C N₂</td>
<td>30' 950°C N₂</td>
<td>42' 950°C N₂, O₂ 140.40</td>
<td>30' 800°C N₂</td>
<td>20' 800°C N₂</td>
</tr>
<tr>
<td>Si₃N₄ CVD</td>
<td>Time Temp. Gases (sccm)</td>
<td>10' 600°C N₂</td>
<td>47' 800°C N₂</td>
<td>22 800°C NH₃, DCS 140.40</td>
<td>40 600°C N₂</td>
<td>10 600°C N₂</td>
</tr>
<tr>
<td>Clean-up Oxidation</td>
<td>Time Temp. Gases</td>
<td>20' 800°C N₂</td>
<td>10' 850°C N₂</td>
<td>10'/14.25'/5' 850°C O₂, Steam, O₂</td>
<td>10' 800°C N₂</td>
<td>20' 800°C N₂</td>
</tr>
<tr>
<td>Field Oxidation</td>
<td>Time Temp. Gases</td>
<td>20' 800°C N₂</td>
<td>35' 1000°C N₂, Low O₂</td>
<td>10'/190'/10' 1000°C O₂, Steam, O₂</td>
<td>30' 800°C N₂</td>
<td>20' 800°C N₂</td>
</tr>
</tbody>
</table>

Table B.2: Furnace Parameters for LOCOS steps. DCS = dichlorosilane

fabrication process is listed in Figure 2.22, Figure 2.23, and Figure 2.23. Specific times and temperatures for the LOCOS processing steps are provided in Table B.2.

### B.4 Gate Oxide Growth

Gate oxidation was performed in dry O₂ at 850°C. Several different strategies were used to get the best possible results from the gate oxidation. Most importantly, all chemicals used in the diffusion clean were fresh. Second, the oxidation was performed early in the morning following an overnight trichloroethane (TCA) clean of the tube. Finally, Tylan-2 was used because that is the tube historically used for BiCMOS processing and an attempt is made to restrict its use to the cleanest processing steps.
<table>
<thead>
<tr>
<th>Wafer Purpose</th>
<th>time (mm:ss)</th>
<th>thickness (nm)</th>
<th>Wafer numbers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Calibration run 1</td>
<td>15:00</td>
<td>5.38</td>
<td>T1.T2</td>
</tr>
<tr>
<td>Calibration run 2</td>
<td>20:00</td>
<td>6.23</td>
<td>T3.T4</td>
</tr>
<tr>
<td>Calibration run 3</td>
<td>30:00</td>
<td>7.67</td>
<td>T1.T2</td>
</tr>
<tr>
<td>Calibration run 4</td>
<td>45:00</td>
<td>9.26</td>
<td>T3.T4</td>
</tr>
<tr>
<td>Gate Oxide</td>
<td>29:07</td>
<td>7.50</td>
<td>T1.30.36.38.39.46-49</td>
</tr>
<tr>
<td>Gate Oxide</td>
<td>52:01</td>
<td>10.46</td>
<td>32-35,40,42-25.T2</td>
</tr>
</tbody>
</table>

Table B.3: Gate oxidation growth conditions

The times for growing gate oxides were determined experimentally. A series of oxidations were performed on monitor wafers and the thickness of the resulting oxide film was measured with a Gaertner ellipsometer. The natural log of the resulting thickness was plotted against the growth time and a least squares fit was performed. The experimental time–thickness relationship obtained is

\[ x = 1.41t^{0.5} \text{nm} \quad \text{Pearson's } r: 0.997. \quad (B.1) \]

where \( t \) is the growth time in minutes and \( x \) is the measured oxide thickness in nm.

Calibration run data and gate oxidation times are presented in table B.3.
APPENDIX C

APPENDIX A-PHASE SHIFT CIRCUITRY

The phase inverting circuit is a derivative of a 360° phase shift circuit presented in National Semiconductor application note AN-32 [85]. It works as follows [86]. A resistor, $R_3$, and capacitor, $C_1$, are connected in series. The AC voltage applied to the sample is applied to the resistor and is inverted. The inverted AC voltage is applied to capacitor $C_1$ (Figure C.1). The output of the divider is phase shifted with respect to the input AC voltage; it is inserted into a unity gain voltage follower that prevents loading of the phase shift network. This is the first stage. The second stage is nearly identical. However, a variable resistor, $R_6$, in the network allows the phase to be adjusted and the buffer amplifier of the second stage has a gain of two. The output of the buffer amp goes to a potentiometer, $R_9$, used as a voltage divider in series with a resistor, $R_{10}$. This allows independent control of the signal amplitude between 0.3 and 2 times the input AC voltage. Finally, a voltage follower is used to buffer the phase shift circuit from the external circuitry.

The inverting and buffer amps in the first and second stage are each 1/4 of a LF347 quad op-amp from National Semiconductor. The output buffer is 1/2 of a LF412 dual op-amp from National Semiconductor. Power to the circuitry is from a
Figure C.1: The phase shift circuit is two stages of a phase splitter followed by a resistor and capacitor voltage divider. The output of each stage is buffered by a voltage follower. The phase shift introduced in the first stage is fixed while the phase shift introduced in the second stage is variable. The output buffer of the second stage is a fixed gain (2) non-inverting amplifier. The amplitude of the phase shifted voltage is controlled by using a potentiometer as a voltage divider. This is buffered by a unity gain voltage follower.

homebuilt ±15V supply. The variable resistors, $R_6$ and $R_9$ in Figure C.1, are both wirewound potentiometers. Component values are presented in Table C.
Table C.1: The component values used in phase shift circuitry are listed here. The component numbers refer to Figure C.1.


[77] aFM tips are CSC-11 coated with TiN. (unpublished).


