Development and Characterization of Low-Disorder Metal-Oxide-Silicon Quantum Dot Devices

Jin-Sung Kim

A Dissertation
Presented to the Faculty of Princeton University in Candidacy for the Degree of Doctor of Philosophy

Recommended for Acceptance by the Department of Electrical Engineering Adviser: Stephen A. Lyon

November 2018
Abstract

Spins confined in Metal-Oxide-Silicon (MOS) quantum dot devices are promising qubits in a quantum processor, demonstrating long coherence times, a large valley splitting, and coherent interactions with donor qubits. Furthermore, the mature fabrication infrastructure of the CMOS industry offers a tantalizing roadmap towards scaling to extremely large quantum systems on a single silicon chip. Despite the incredible advances in materials and fabrication developed by the classical CMOS industry and continued by the silicon quantum electronics community, many challenges remain in building a silicon quantum processor. One of the biggest challenges impeding the scaling of large quantum dot systems is the presence of disorder and shallow electron traps at the Si/SiO$_2$ interface. The purpose of this thesis is to illuminate mechanisms of disorder at the Si/SiO$_2$ interface relevant to quantum dot devices, develop a framework for understanding and quantifying shallow electron traps in terms of electron spin resonance and transport measurements, and to pave a path forward for scaling MOS quantum dot devices.

This thesis is organized into three sections: first, we develop a fabrication process in silicon MOSFETs yielding a low-disorder Si/SiO$_2$ interface in order to leverage this process as a platform for fabricating low-disorder quantum dot devices. One of the challenges in working with silicon oxide is that high energy processes and mobile ionic contamination during the fabrication process can create electron traps and disorder at the Si/SiO$_2$ interface. Using the low temperature (4.2 K) electron mobility as a proxy for the Si/SiO$_2$ interface quality, we study the effect of various processing parameters on the Si/SiO$_2$ interface disorder and ultimately arrive at a fully optimized process yielding the highest reported mobility (23,000 cm$^2$/Vs) thin-oxide ($\leq$30 nm) silicon MOSFET.

Secondly, we study the annealing of shallow electron traps created by electron-beam (e-beam) lithography. E-beam lithography is a necessary tool in defining nano-
scale electrostatic gates which define the quantum dot potential, but the high-energy electrons and photons created in the process create electron traps at the interface. We directly probe shallow electron traps using electron spin resonance (ESR) and demonstrate that 1) a standard forming gas anneal is sufficient to passivate electron traps created by the e-beam exposure, and 2) that our lowest temperature ESR measurements agree with transport measurements of the devices’ percolation threshold, demonstrating agreement between two independent methods of characterizing the Si/SiO$_2$ interface.

Finally, leveraging the above process optimizations, we fabricate and characterize a low-disorder double quantum dot device. We demonstrate agreement between the dots’ charging energy and lithographic size, concluding that our dots are lithographically defined and not dominated by random disorder. Charge sensing measurements indicate regular quantum dot transitions over a wide parameter range down to the single electron regime, with evidence of few defects in the vicinity of the quantum dots, and the controllable formation of a quantum double dot. Noise spectroscopy measurements of the dot indicate a $1/f$ like power spectral density that is comparable in magnitude to other Si quantum dot devices measured at 300 mK. Finally, magneto-spectroscopy measurements of the first and second electron transitions yield a valley splitting of 110±26 µeV, large enough to support high-fidelity spin selective operations.

With this work, we demonstrate a method of fabricating low-disorder, high-mobility silicon MOSFETs, a framework for studying disorder in quantum dot devices in the low-electron density regime, and a promising platform for MOS qubits in a low-disorder quantum dot device architecture.
Acknowledgements

These past six years at Princeton have been a privilege and I have been graced by the presence of many people around me who have enabled me to not only survive but thrive during my PhD studies. Indeed the successes I have experienced in graduate school, both personal and academic, have only been made possible by standing on the shoulders of giants whom I would like to acknowledge below.

First and foremost I would like to thank my advisor Steve Lyon who first hired me as a summer researcher the summer between my senior year and MEng at Cornell. Steve opened the door to my graduate PhD career and with his recommendation I returned to Princeton a year later as a fulltime PhD student. As an advisor he was constant source of knowledge and ideas, and taught me to be a careful and thorough experimentalist. His infectious love of science and unwavering optimism have inspired me to continue in a career of science after Princeton.

I would also like to thank all of the other professors in Electrical Engineering who, through classes and many hallway conversations, have provided me with a solid foundation of knowledge. I would like to thank especially Professors Nathalie de Leon and Jim Sturm for reading my thesis and providing me with many helpful comments, and Professors Andrew Houck and Mansour Shayegan for being on my thesis committee, reprising their roles as my generals committee.

I would like to thank the members of the Lyon lab past and present (Alexei Tyryshkin, Maika Takita, Ryan Jock, Evan Petersen, Brendon Rose, Anthony Sigillito, Abraham Asfaw, Kyle Castoria, and Ethan Kleinbaum) for being excellent company and for reminding me that research isn’t meant to be easy but also reminding me that everyone is in the same boat. I would like to especially thank Ryan for mentoring me when I first joined the group and for continuing to provide me with valuable research and life advice. Alexei for being an invaluable source of knowledge
and for always being willing and available to help. Brendon for listening to my gripes. Ethan for also listening to my gripes and helping to restart Thursday liquid analysis.

I want to thank the many friends at Princeton I have made over the years. The original lunch group: Yasmin Afsar, Ken Nagamatsu, Matt Chang, Sunil Pandey, Omer Malik, Harvey Cheng, Daniel Jiang, Andrew Kim, Sachin Ravi. Volleyball friends: Josephine Lembong, Jenny Yu, Manuel Mueller, Renaud Gueroult, Carsten Milsman, Jonathan Lin, Patrick Signoret, Vlad Feinberg, Ting Chen and all the members of the Princeton Men’s Club Volleyball team. Basketball friends: Joe Durante, Eric Blow, Fermi Ma, Bruce Allen, Bobby Lyon, Bobby Hristov, Winston Chou, Peng Chen, Rafael Parente. Other great friends: Mitch Nahmias, Sara Chuang, Brandy Briones and Weston Fleming.

I thank Whitman College for being my home for four years and the college staff, especially Rebecca Grave-Bayazitoglu and Kristin Frasier. Additionally, Amir Roknabadi for being a great roommate and friend.

I want to thank the PRISM cleanroom staff past and present, especially: Pat Watson, Conrad Silvestre, Joe Palmer, Yong Sun, Eric Mills, Bert Harrop. The electrical engineering administrative staff, especially: Barbara Fruhling, Jessica Johnson, Colleen Conrad, Sara McGovern, and Linda Dreher.

I am indebted to Susanne Killian whose many appointments with me at Career Services were indispensable throughout my career search.

A special thanks to my significant other, Liz Davison, for sharing my highs and lows during this time, for always believing in me and for always supporting me despite all of my neuroticism. You make me a better, more complete, more thoughtful person. I cannot imagine what graduate school would have been without meeting you.

Finally I thank my family. Without them, none of my successes would have been possible. I thank my siblings Yei-Sung (“Noona”), Hyun-Sung (“Brother”), and Mi-Sung (“Sister”) who all had a hand in raising and educating me.
Most of all, I thank my parents who instilled in me the values of education, hardwork, and humility. Throughout my PhD, they kept me grounded and provided me with perspective, reminding me that when problems in your research are the biggest problems in your life, you’re living a good life. Their hardwork and sacrifice have enabled me to be where I am now. This thesis is dedicated to them and belongs to their legacy.
To my parents.
Contents

Abstract ................................................................. iii
Acknowledgements ....................................................... v
List of Tables .......................................................... xiii
List of Figures .......................................................... xiv

1 Introduction ......................................................... 1
   1.1 Overview of Silicon Quantum Dot Devices ......................... 3
       1.1.1 Background .............................................. 3
       1.1.2 Introduction to Quantum Dot Physics ...................... 6
   1.2 MOS vs. Si/SiGe .................................................. 14
       1.2.1 Integration with donors .................................. 14
       1.2.2 Valley Splitting ......................................... 16
       1.2.3 Interface Disorder ....................................... 17
   1.3 Outline of Thesis ................................................ 19

2 Introduction to the MOS System .................................. 21
   2.1 The MOSFET .................................................... 21
   2.2 MOSFET Operation .............................................. 22
       2.2.1 Threshold Voltage ....................................... 23
       2.2.2 Capacitance ............................................. 25
       2.2.3 Current-Voltage (I-V) Characteristics .................... 30
4.4.2 Percolation Threshold ................................................. 82
4.4.3 Electron Spin Resonance Measurements ......................... 85

4.5 Conclusions .............................................................. 88

5 Metal-Oxide-Semiconductor Double Quantum Dots .................. 90

5.1 Introduction .................................................................. 91
5.2 Device Fabrication ....................................................... 93
5.3 Experimental Setup ...................................................... 96
5.4 Tuning a Quantum Dot .................................................. 100
5.5 Bias Spectroscopy ........................................................ 103
5.6 Charge Sensing Measurements ....................................... 104
  5.6.1 Quantum Dot Transitions ......................................... 105
  5.6.2 “Phantom Dot” Transitions ....................................... 108
  5.6.3 Defect Detection .................................................... 109
  5.6.4 Double Dot .......................................................... 110
5.7 1/f Charge Noise ......................................................... 110
5.8 Valley Splitting ........................................................... 112
5.9 Conclusions ............................................................... 115

6 Conclusion ...................................................................... 116

6.1 Future Work ............................................................... 119

A Device Fabrication Details ................................................. 122

A.1 Notes on Contamination ................................................. 122
A.2 MOSFET Fabrication .................................................... 125
  A.2.1 Detailed Fabrication Process ...................................... 125
  A.2.2 Notes on the Geometry and Placement of the Source/Drain Contacts in Large-Area MOSFETs ................................................. 127
  A.2.3 Common Failure Modes and Troubleshooting ................. 128
<table>
<thead>
<tr>
<th>A.3</th>
<th>Quantum Dot Fabrication</th>
<th>130</th>
</tr>
</thead>
<tbody>
<tr>
<td>A.3.1</td>
<td>Detailed Fabrication Process</td>
<td>130</td>
</tr>
<tr>
<td>A.4</td>
<td>Recipes</td>
<td>133</td>
</tr>
<tr>
<td>A.4.1</td>
<td>Photolithography</td>
<td>133</td>
</tr>
<tr>
<td>A.4.2</td>
<td>Electron-beam Lithography</td>
<td>135</td>
</tr>
<tr>
<td>A.4.3</td>
<td>Solvent Clean</td>
<td>136</td>
</tr>
<tr>
<td>A.4.4</td>
<td>RCA Clean</td>
<td>136</td>
</tr>
<tr>
<td>A.4.5</td>
<td>Piranha Clean</td>
<td>137</td>
</tr>
<tr>
<td>A.4.6</td>
<td>Resist Stripping</td>
<td>137</td>
</tr>
<tr>
<td>A.4.7</td>
<td>Wet Chemical Etching of Polysilicon</td>
<td>138</td>
</tr>
<tr>
<td>A.4.8</td>
<td>Savannah ALD Aluminum Oxide Growth Conditions</td>
<td>139</td>
</tr>
<tr>
<td>A.4.9</td>
<td>Chlorine Cleaning of Furnace Tubes</td>
<td>140</td>
</tr>
</tbody>
</table>

**Bibliography**

141
List of Tables

3.1 Table of electron mobility from polysilicon etch parameters . . . . . . 59
3.2 Table of electron mobility from before and after bias temperature stress 60
3.3 Table of electron mobility of samples annealed with and without alu-
  minum . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 65
List of Figures

1.1 Schematic of quantum dot device cross section and ideal potential profile 7
1.2 Schematic of quantum dot electrochemical potential 12

2.1 MOSFET schematic in the off configuration 22
2.2 MOSFET schematic in the on configuration 25
2.3 Simple picture of the MOS capacitance 27
2.4 MOS CV curve 29
2.5 MOSFET Drain Current Family of Curves 30
2.6 Four types of oxide defects 34
2.7 Threshold voltage shift in MOS CV curve 35
2.8 The Deal-Grove triangle 39
2.9 High frequency MOS CV curve compared to low frequency CV curve 41

3.1 Hall bar geometry and optical photograph 48
3.2 Schematic of base MOSFET fabrication process 50
3.3 Hall mobility measurement setup 51
3.4 Electron density measured from the Hall voltage 54
3.5 4.2 K Electron mobility for base process 56
3.6 4.2 K Electron mobility comparison for various polysilicon etch parameters 58
3.7 Mobility comparison of bias temperature stressed 61
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.8</td>
<td>Electron density comparison of bias temperature stressed</td>
<td>62</td>
</tr>
<tr>
<td>3.9</td>
<td>Mobility comparison of aluminum annealing</td>
<td>64</td>
</tr>
<tr>
<td>4.1</td>
<td>Electron beam damage to the Si/SiO(_2) interface</td>
<td>71</td>
</tr>
<tr>
<td>4.2</td>
<td>Large area ESR MOSFET and Hall bar</td>
<td>73</td>
</tr>
<tr>
<td>4.3</td>
<td>Diagram of process splits for e-beam irradiation experiments</td>
<td>74</td>
</tr>
<tr>
<td>4.4</td>
<td>Schematic of the (^3)He cryostat</td>
<td>77</td>
</tr>
<tr>
<td>4.5</td>
<td>Schematic of the basic ESR setup</td>
<td>79</td>
</tr>
<tr>
<td>4.6</td>
<td>Temperature dependence of the low temperature mobility of samples A and B</td>
<td>80</td>
</tr>
<tr>
<td>4.7</td>
<td>Mobility comparison of samples A, B, and C</td>
<td>81</td>
</tr>
<tr>
<td>4.8</td>
<td>Cartoon of disorder in a 2D energy landscape</td>
<td>82</td>
</tr>
<tr>
<td>4.9</td>
<td>Percolation threshold plotted for samples A and B as a function of temperature</td>
<td>84</td>
</tr>
<tr>
<td>4.10</td>
<td>Comparison of unpaired spins for dark and post-illuminated ESR MOS-FET</td>
<td>86</td>
</tr>
<tr>
<td>4.11</td>
<td>ESR measured shallow trap densities for samples A and B as a function of temperature</td>
<td>88</td>
</tr>
<tr>
<td>5.1</td>
<td>Fabrication flow diagram of double quantum dot device</td>
<td>93</td>
</tr>
<tr>
<td>5.2</td>
<td>False colored SEM image of double quantum dot device with cross-section</td>
<td>94</td>
</tr>
<tr>
<td>5.3</td>
<td>Photograph of quantum dot device bonded to PCB</td>
<td>95</td>
</tr>
<tr>
<td>5.4</td>
<td>Mobility of SiO(_2)/ALD/Al gate stack</td>
<td>96</td>
</tr>
<tr>
<td>5.5</td>
<td>Block diagram of instrumentation set up for quantum dot transport measurements</td>
<td>97</td>
</tr>
<tr>
<td>5.6</td>
<td>Photograph of instrumentation set up for quantum dot transport measure-ments</td>
<td>98</td>
</tr>
<tr>
<td>5.7</td>
<td>Schematic of the breakout box used for QD measurements</td>
<td>99</td>
</tr>
</tbody>
</table>
5.8 Pinch-off voltages for double quantum dot device .................................. 101
5.9 2D conductance scan of tunnel barriers ...................................................... 101
5.10 Measured Coulomb blockade oscillations .................................................. 102
5.11 Measured Coulomb blockade diamonds for left and right quantum dots .......... 103
5.12 Charge sensor drain current ........................................................................ 105
5.13 Charge stability diagrams for left and right quantum dot .............................. 106
5.14 Bias spectroscopy plot of the first excited state ........................................... 108
5.15 Coulomb blockade drain current ................................................................. 111
5.16 Quantum dot 1/f charge noise ................................................................... 111
5.17 Measured MOS quantum dot valley splitting ............................................. 113

A.1 Iterations of source/drain gate configurations for large-area MOSFETs ........ 128
Chapter 1

Introduction

Over the past two decades, a tremendous amount of progress has been made towards building a quantum computer in a myriad of different architectures and materials systems, including trapped ions \cite{1, 2}, superconducting circuits \cite{3}, and electrons on helium \cite{4}. One of the first architectures proposed were electron spins confined in coupled semiconductor quantum dots in Loss and DiVincenzo’s seminal 1999 paper \cite{5}. Quantum dots, or artificial atoms, are artificial electrostatic potentials, often times defined by electrostatic gates, which define a three dimensional confinement well small enough to probe discrete charge, spin, and orbital states. This proposal was motivated by three attractive features of the semiconductor quantum dot \cite{5}. First, the rapid scaling of semiconductor microelectronics to smaller and denser features within a single semiconductor chip offers a tantalizing pathway forward for scaling to the billions of quantum bits \cite{6} (or qubits) necessary for a universal quantum computer. Second, the electron spin presents a well-defined qubit, as a true two-level system with no other states for information to leak to. Third, the exchange interaction between two electrons in adjacent quantum dots offers a controllable way of creating entangled states.
Later, DiVincenzo codified the requirements of a universal quantum computer in the now famous “DiVincenzo Criteria:” \[7\]

1. A scalable physical system with well characterized qubits

2. The ability to initialize the state of qubits

3. Coherence times long compared to the gate operation time

4. The ability to perform one and two-qubit gates

5. The ability to read-out the qubit state

At the time of Loss and DiVincenzo’s proposal, the semiconductor quantum dot substrate of choice was AlGaAs/GaAs heterostructures grown by molecular-beam epitaxy (MBE). These substrates are able to achieve extremely high quality, nearly defect-free, ultra-high mobility interfaces in which to confine electrons. Indeed, much of the foundational work on quantum dot spin systems for quantum computing purposes was developed in AlGaAs/GaAs substrates, including the development of highly symmetric lithographically defined quantum dots \[8\], qubit initialization \[9\], coherent single qubit manipulation \[10, 11\], and single-shot spin readout \[9\].

Despite the many advances made towards spin qubits in GaAs quantum dots, an intrinsic obstacle limits GaAs spin qubits for real quantum computers: short coherence times caused by the large percentage of nuclear spin isotopes \[11\]. This obstacle motivated the need for a new material substrate for quantum dot spin qubits, a material with a low natural percentage of nuclear spin isotopes yielding long coherence times, a material with decades of research and development behind it and a mature fabrication infrastructure in place: silicon.
1.1 Overview of Silicon Quantum Dot Devices

Silicon provides the backbone of modern classical computing and may be the most technologically advanced material system today, having benefited from decades of advancements thanks to the CMOS industry. Fortuitously, silicon’s intrinsic materials properties also make it a promising host for spin qubits in a quantum computer. However, silicon also offers new challenges in quantum dot devices in the form of interfacial disorder and multi-valley physics [12, 13].

1.1.1 Background

Spins in silicon have demonstrated very long coherence times ($T_2$), i.e. the timescale which characterizes how long a quantum state may be held in a superposition before the information leaks out to the environment [14, 15]. These long coherence times are attributed to silicon’s low natural spin-orbit coupling and a low natural percentage of nuclear spin isotopes (5% $^{29}$Si). In highly isotopically enriched silicon crystals, where the residual $^{29}$Si has been reduced to 50 ppm, spin coherences have exceeded 10 seconds [14].

In a gate defined quantum dot device, electrons are confined within a quantum well forming a two dimensional electron gas (2DEG) where lithographically defined electrostatic gates then form a lateral confinement potential. Within silicon materials systems, quantum dots are typically fabricated on one of two types of heterostructures, the metal-oxide-semiconductor (MOS) stack and Si/SiGe quantum wells. In MOS structures, electrons are confined at the Si/SiO$_2$ interface in a triangular potential well. Here, an amorphous oxide layer is grown from a crystalline silicon substrate resulting in a crystalline-amorphous interface. The SiO$_2$ provides a large energy barrier which insulates the 2DEG from the overlying gates. In contrast, in a Si/SiGe heterostructure, electrons are confined within a 10 nm Si quantum well, sandwiched
between two Si_{0.7}Ge_{0.3} buffer layers. Si/SiGe quantum wells are typically grown by chemical vapor deposition or molecular-beam epitaxy, so the Si quantum well can be atomically sharp, albeit with atomic steps at the interface.

The essential requirement in quantum dot qubits is to controllably confine a single electron within a quantum dot. Despite the mature fabrication infrastructure developed by the CMOS industry, now fabricating transistors at the 7 nm node level, the development of silicon quantum dot devices has emerged only relatively recently, where the first controllable confinement of electrons demonstrating regular Coulomb blockade oscillations in a silicon quantum dot was demonstrated in 1999 [16]. One of the main difficulties in achieving single electron occupancy is the presence of defects within the quantum well in which the quantum dot is defined [17, 18, 19, 20, 13, 21]. In addition, the defects that affect silicon quantum devices are distinct from the ones studied and optimized for classical silicon devices because quantum devices must operate at cryogenic temperatures. Therefore, the ability to simply pattern nanoscale features, like in a CMOS chip, does not directly translate to the ability to fabricate quantum dot devices.

In both MOS and Si/SiGe quantum dots, the requirements for a functional qubit remain the same: the ability to confine a single (or few) electron in the quantum dot, and the ability to do spin selective operations. Single electron occupation was first achieved in Si/SiGe quantum dots in 2007 [22], and shortly after it was demonstrated in MOS quantum dots [23]. In the Si/SiGe case, the authors utilized a quantum point contact (QPC) adjacent to the quantum dot in order to detect the tunneling of the last electron out of the quantum dot. In the MOS work, the authors relied on observing the tunneling of electrons through the quantum dot itself without using a charge sensor. Without an integrated charge sensor, however, there is the danger that the last observed tunneling events do not correspond to the last electron transitions. Two years later, the same group implemented a charge sensor adjacent to their quantum
dot and explicitly demonstrated the occupation of their MOS quantum dot down to the single-electron regime \[17\].

Since the achievement of single-electron occupation, a variety of spin selective operations of a single spin in silicon quantum dot systems have been achieved. Pauli blockade, the action of blocking an electron into a quantum dot based on its spin state, was achieved in both MOS and SiGe double quantum dot systems in 2010 and 2011 \[24, 25\]. The use of Pauli blockade is an effective way of initializing the spin state of a qubit. It should be noted that the MOS double dot was not in the single electron regime, rather a regime with approximately ten electrons in each dot, whereas in the SiGe work, the double dot was in the true single electron regime.

Parallel to this effort confining single electrons in a QD, a group demonstrated the single-shot read-out of an electron spin bound to a \(^{31}\)P donor utilizing a nearby quantum dot charge sensor \[19\]. This work circumvents the need for confining a single electron within a quantum dot for use as a qubit by utilizing a single electron bound to a phosphorus donor. In this work, several phosphorus donors were implanted adjacent to a MOS quantum dot device. Depending on the spin state of the electron bound to the donor, the electron could selectively tunnel into the quantum dot, resulting in a measurable change in current through the quantum dot. This work demonstrates the deterministic initialization and readout of spin qubits in silicon.

Having demonstrated Pauli spin blockade in Si QDs, various groups have demonstrated a myriad of spin selective operations. In 2012, coherent singlet-triplet rotations were demonstrated in a SiGe double quantum dot \[26\]. In MOS systems, the coherent singlet-triplet rotations of an electron were demonstrated mediated by the hyperfine interaction of a nearby donor \[27\].

Very recently, groups in both MOS and SiGe have demonstrated a two qubit CNOT operation between electron spins in adjacent quantum dots. In the MOS group, the authors utilize a microwave antenna to selectively drive qubit rotations,
where the single qubit resonant frequency can be tuned by the exchange interaction between the two qubits [28]. In the SiGe group, the authors utilize a cobalt micromagnet to apply a magnetic field gradient and modulate the electric field applied to one of the qubits, driving the qubit rotation through an effective magnetic field modulation [29].

Moving forward, one of the outstanding challenges in MOS devices is to uniformly control and confine single electrons in quantum dots in the absence of defects and disorder. Overcoming this challenge requires careful study of materials and processing while understanding and minimizing the relevant defects for these devices.

1.1.2 Introduction to Quantum Dot Physics

Quantum dots are artificially defined potential wells which can be filled with individual electrons (or holes) [30, 31]. Metallic gates are patterned on the surface of the substrate which define the dot potential, as shown in Figure 1.1. The potential well in which the electrons are confined is capacitively coupled to electron reservoirs via tunnel barriers. If the dot is connected to two reservoirs which are connected to ohmic contacts, the dot is sometimes referred to as a single electron transistor (SET) where the two reservoirs can be biased to form a source and drain. In most recent quantum dot designs, the potentials of the dot, the tunnel barriers, and the reservoirs can be independently adjusted by individual electrostatic gates [32, 27, 28, 33].

At cold enough temperatures and with tight enough confinement of the dot, one can populate the dot with individual electrons such that the charge state is restricted to a definite number of electrons \( N \). The confinement can be characterized by the charging energy of the dot \( E_C \), that is the energy required to add one more electron to the dot. Thus the relevant energy scale and temperature scale to confine individual electrons is \( k_B T \ll E_C \). For reference, \( k_B T \) at 1 K is 87 µeV.
Quantum Confinement

If we treat the dot as a metallic disc, the charging energy can be well approximated by the self capacitance of this metallic disc in a dielectric medium through the relation [31]:

\[ E_C = \frac{e^2}{C_{\text{disc}}} \]  \hspace{1cm} (1.1)

\[ C_{\text{disc}} = 8\varepsilon_r\varepsilon_0 R \]  \hspace{1cm} (1.2)

Here, \( e \) is the electron charge, \( \varepsilon_r \) is the relative permittivity of the medium where the dot is defined, \( \varepsilon_0 \) is the permittivity of free space, and \( R \) is the radius of the dot. Therefore, a smaller dot leads to greater confinement which leads to larger charging
energies. For a 35 nm radius dot at the Si/SiO$_2$ interface ($\epsilon_r = \frac{1}{2}(\epsilon_{Si} + \epsilon_{SiO_2})$), $E_C$ is about 8 meV.

Wrapped into this approximation is what is known as the Constant Interaction (CI) model [31]. The CI model parameterizes all of the Coulomb interactions of the electrons within the dot and folds them into a single constant capacitance, $C$, where $C$ is the sum of all capacitances between the dot and various electrodes surrounding the dot. To a good approximation, $C$ can be estimated by equation 1.2. That is:

$$C = \sum C_i = C_{Self} + C_{Gate} + C_{Source} + C_{Drain} \ldots \approx 8\epsilon_r\epsilon_0 R$$ (1.3)

Additionally, the CI model assumes that the single-particle energy spectra are independent of the electron-electron interactions within the dot [31]. Therefore, the $N$th electron added to the dot occupies the $N$th orbital state, neglecting spin and valley degeneracy.

**Tunnel Barriers**

In general, lithographically defined quantum dots are tunnel coupled to an electron reservoir on either side of the dot in order to tune the electron population in the dot and to perform transport measurements through the dot. These tunnel barriers must be opaque enough such that quantum fluctuations in the number of electrons within the dot is much less than one over the course of the measurement. In other words, for good confinement, the tunnel rate must be slow enough that electron transitions in and out of the dot are not lifetime broadened through Heisenberg uncertainty.

We can calculate a lower bound for the resistance of the tunnel barrier to avoid lifetime broadening the electron transitions [31]. An estimate for the time it takes to charge or discharge the dot can be calculated by the RC time constant of the dot: $\Delta t = R_tC$ where $R_t$ is the tunnel barrier resistance and $C$ is again the dot
capacitance. Heisenberg uncertainty requires $\Delta E \Delta t > h$. Plugging in the charging energy and the RC time constant:

$$\left(\frac{e^2}{C}\right) (R_t C) > h$$

$$R_t > \frac{h}{e^2}$$

We see that the tunnel barrier resistance must be greater than resistance quantum, $h/e^2 = 25.813 \text{ kΩ}$.

In order to probe discrete charges on the dot then, the two conditions that must be met are:

$$E_C \gg k_B T$$

$$R_t \gg \frac{h}{e^2}$$

**The Energy of a Quantum Dot**

The total energy of the dot $U(N)$ can be calculated by the following [30]:

$$U(N) = \frac{-|e|(N - N_0) + \sum_i C_i V_i^2}{2C} + \sum_{n=1}^{N} E_n(B)$$

Here, $N_0$ is charge on the dot induced from background positive charge present in the substrate. The first term in expression [1.8] is simply the energy stored in the dot capacitance with induced charge on the dot from $i$ electrodes around the dot and from the positive background charge. The second term is the sum of the orbital energies $E_n(B)$ as electrons occupy higher single-electron orbital states as more electrons are added to the dot. These orbital energies are in general functions of the magnetic field $B$. 
The orbital energy quantization in the dot is commonly estimated simply by treating the electrons as laterally confined in a 2D box which is defined by the geometry of the electrostatic gates [32, 34]. For a 2D box, $\Delta E_{2e+1} = \frac{3\hbar^2 x^2}{2m^* L^2}$, where $m^*$ is the effective mass of the electron and $L$ is the length of the box. In silicon, the effective mass $m^*$ is $m_l = 0.92m_0$ in the longitudinal direction and $m_t = 0.19m_0$ in the transverse direction. Because electrons confined in a silicon quantum dot preferentially occupy the valley state perpendicular to the quantum well, the appropriate effective mass to utilize is the transverse effective mass. For an electron confined in a $L = 70$ nm box, $\Delta E$ is 1.2 meV.

It is also common to treat the lateral confinement as a 2D isotropic parabolic well $V(x, y) = K/2(x^2 + y^2)$ where the single electron energy eigenvalues are [35]:

$$E_{n_x, n_y} = \hbar (K/m_t)^{1/2} (n_x + n_y + 1) \quad (1.9)$$

$$n_x, n_y = 0, 1, 2, ... \quad (1.10)$$

Here $K$ characterizes the strength of the parabolic potential and is related to the characteristic length of the well $l$ by $l = 2\hbar ^{1/2} (Km_t)^{-1/4}$.

In the perpendicular direction, confinement is achieved by confining the electrons to the quantum well interface in a triangular potential well. The orbital energy within a triangular well is given by $E_N = -\left(\frac{e^2 \hbar^2}{2m}a_n\right)^{1/3}$. Here $E$ is the electric field and $a_n$ is the $n$th root of the Airy function and is a negative value. The first excited state of the triangular potential well is about 10 meV above the ground state, so in general it is safe to assume that electrons reside in the ground state of this potential. Therefore the orbital excitation spectra is dominated by the lateral confinement potential.
The Electrochemical Potential

The most convenient quantity for describing the state of the quantum dot is the electrochemical potential, $\mu(N)$ [30]. By definition, the electrochemical potential is $\frac{dU(N)}{dN} = U(N) - U(N-1)$. This yields:

$$\mu(N) = (N - N_0 - 1/2)E_C - \frac{E_C}{|e|} \sum_i C_i V_i + E_N$$  \hspace{1cm} (1.11)

From here it is straightforward to calculate the addition energy:

$$E_{add}(N) = \mu(N+1) - \mu(N) = E_C + \Delta E$$ \hspace{1cm} (1.12)

Where $\Delta E$ is the difference the the orbital energies of the $(N+1)\text{th}$ and $N\text{th}$ electrons. So we see that the addition energy is indeed simply the charging energy with a small correction for the orbital energy.

Coulomb Blockade

The alignment of electrochemical potential of the dot, $\mu_{dot}$, relative to the electrochemical potentials of the source and drain reservoirs is the crux of the operation of the single electron transistor. For $k_B T \ll E_C$, electrons can only tunnel in or out of the dot if the $\mu_{dot}$ is aligned within $\sim k_B T$ of $\mu_{source}$ or $\mu_{drain}$.

If a small bias is applied across the source drain reservoirs, $V_{SD} = V_S - V_D$ and $|e|V_{SD} \ll E_C$, a bias “window” is opened such that if $\mu_{dot}$ resides within this energy window $eV_{SD}$, individual electrons can tunnel from one reservoir into the dot and then into the other reservoir (Figure 1.2). In this configuration a “charge degeneracy” exists in the dot such that the number of electrons that are allowed on the dot are $N$ and $N+1$. If $\mu_{dot}$ resides outside of this bias window, electron tunneling is forbidden as there is no available state in the dot within $k_B T$ of electrons in source/drain reservoirs.
Figure 1.2: Schematic of the energy structure of a quantum dot taken from reference [30]. a) Alignment of the dot electrochemical potential such that $\mu_{\text{dot}}$ resides outside of the bias window defined by $\mu_S$ and $\mu_D$. The dot is Coulomb blockaded in this configuration. b) Alignment of $\mu_{\text{dot}}$ within the bias window. c) Current measured through the dot scanning the gate voltage coupled to the dot. Each time $\mu_{\text{dot}}$ aligns within $\mu_S$ and $\mu_D$, a spike in current is observed.

for electrons to tunnel into. Therefore within this region, the electron number within the dot is fixed to $N$. This condition is known as Coulomb blockade.

If we now scan the dot gate voltage $V_g$, which couples to the dot’s electrochemical potential, each time $\mu_{\text{dot}}$ aligns with the bias window of $\mu_{\text{Source}}$ and $\mu_{\text{Drain}}$ we observe spikes of current known as Coulomb blockade peaks (Figure 1.2 (c)). These repeating Coulomb blockade peaks are known as Coulomb oscillations. Looking at a region on the left side of a Coulomb blockade peak and comparing it to the region to the right of the Coulomb blockade peak we know that the electron number within the dot has increased by exactly one.
In the regime of $k_B T \ll E_C$, the line shape of the Coulomb blockade peak now depends on whether $k_B T$ is greater than or less than the difference in orbital energies $\Delta E$ within the dot. For $k_B T \gg \Delta E$, many orbital states are thermally accessible when electrons tunnel through the dot. This is known as the classical or metallic Coulomb blockade regime and the lineshape of the conductance $G$ of a Coulomb blockade peak is given by \[37\]:

$$G/G_\infty \approx \frac{1}{2} \cosh^{-2} \left( \frac{\alpha |V_{g,\text{res}} - V_g| e}{2.5 k_B T} \right)$$

(1.13)

In the above expression, $G_\infty$ is the conductance through both tunnel barriers, $1/G_\infty = 1/G_{\text{left}} + 1/G_{\text{right}}$. $\alpha$ is the known as the lever arm, a measure of the capacitive coupling of the gate to the dot, $\alpha = C_{\text{gate}}/C$. $V_{g,\text{res}}$ is the gate voltage at which the dot electrochemical potential is aligned midway between the source and drain electrochemical potentials.

If $k_B T$ is much less than the difference in orbital energies $k_B T \ll \Delta E$, only a single orbital state will contribute to conduction through the dot at a time. This regime is known as the quantum Coulomb blockade regime and the conductance lineshape is given by \[37\]:

$$\frac{G}{G_\infty} = \frac{\Delta E}{4 k_B T} \cosh^{-2} \left( \frac{\alpha |V_{g,\text{res}} - V_g| e}{2 k_B T} \right)$$

(1.14)

In the quantum Coulomb blockade regime, individual orbital states can be probed so that the dot’s orbital spectrum can be experimentally mapped out.

The functional forms of the lineshape in the classical and quantum Coulomb blockade regimes are very similar. Notice though the temperature dependence of height of the conductance peak in the quantum case and the different effective temperatures. The temperature dependence of the height of the Coulomb blockade peak is a way of distinguishing whether one is operating in the classical or quantum regime.
By fitting the Coulomb blockade peak, many useful quantities can be experimentally extracted. By tracking the linewidth of a Coulomb blockade peak as a function of temperature at high temperature, the lever arm $\alpha$ can be calculated. The lever arm is necessary in converting the gate voltage value to energies within the quantum dot. With the same measurement, one can calculate the electron temperature at low temperature. Often times the electron temperature is significantly higher than the cryostat temperature. If the electrons are fully thermalized with the bath, one expects a linear decrease in the Coulomb blockade linewidth with temperature. Usually the linewidth saturates at a certain temperature and so the true electron temperature corresponds to the temperature where the linewidth saturates [38, 39].

When biased to the edge of a Coulomb blockade peak, the single electron transistor can be used as an extremely sensitive charge sensor of the local electrostatic environment [40, 32, 19]. This is extremely useful when the SET is adjacent to a second quantum dot. In this configuration, the SET can detect the addition of single electrons to the adjacent quantum dot. The SET charge sensor is an indispensable tool in quantum dot experiments.

1.2 MOS vs. Si/SiGe

Within these two subsystems there are three main trade-offs: integration with donors, valley splitting, and disorder.

1.2.1 Integration with donors

In MOS, individual donors can be implanted near the Si/SiO$_2$ interface and can be coupled to a nearby quantum dot [41, 27, 42, 19]. Donor spins have extremely long coherence times due to the weak coupling to the environment [14, 43]. In addition, the combined electronic and nuclear states of donors provide a Hilbert space in which
to perform two-qubit operations, for example using the spin 1/2 nuclear state of a $^{31}$P donor in conjunction with the spin 1/2 electron spin states [43]. These hybrid electron-nuclear spin states allow two two-level systems where the electron spin transition frequency is shifted depending on the spin state of the nuclear spin via the hyper-fine interaction. This naturally allows a 2 qubit CNOT operation utilizing standard ESR and NMR pulses. This enables donors to be used as 2 qubit quantum processors as well as quantum memories.

Recently $^{209}$Bismuth has attracted attention by researchers for its 9/2 spin nucleus [44, 45, 46]. Bismuth’s combined nuclear and electronic spin states allow for so-called clock transitions, spin transitions whose transition frequencies are insensitive to magnetic field variations to first order. These clock transitions have been shown to support coherence times exceeding seconds [44]. Additionally, at every clock transition there are two nearly degenerate transitions which can be individually addressed by microwaves of opposite helicity, creating a 2-qubit subspace useful for computations [46]. This property of Bi donors has motivated several proposals utilizing the electronic-nuclear spin states as physical qubits in a surface code architecture, an error correcting scheme comprised of a 2D array of physical qubits interwoven with stabilizer qubits [47]. In this scheme, Bi donors are implanted in a 2D array and are coupled to one another via electrons confined in nearby quantum dots. Electrons are shuttled from one donor site to another and a SWAP operation is performed between the Bi donor and electron. The computation is then performed on the Bi donor and swapped back to the electron which is then shuttled to the next Bi donor site.

Implanting individual donors in a Si/SiGe substrate is difficult as the donor must be implanted precisely within the $\sim$10 nm silicon quantum well in order for to avoid ionizing the donor [48]. If the donor is implanted instead into one of the Si$_{0.7}$Ge$_{0.3}$ buffer layers sandwiching the Si quantum well, the electron associated with the donor
will prefer to populate the silicon quantum well instead of binding to the donor, achieving a lower ground state energy this way.

Furthermore, even if the donor is implanted within the silicon quantum well of a SiGe substrate, it is difficult to activate the implant, i.e. incorporate the donor atom into the silicon lattice, without damaging the strained silicon quantum well. The usual way to activate implanted donor atoms in a silicon substrate is to anneal the substrate at high temperatures (≥900 C) [49]. However, Si/SiGe quantum wells are typically grown at much lower temperatures (~600 C) in order to lattice match the silicon layer to the SiGe buffer layers [50, 51]. By bringing the substrate above the growth temperature, the strained silicon layer is allowed to relax, thus destroying the high-quality lattice matched Si/SiGe interface. Thus, the thermal budget of the Si/SiGe substrate does not allow for the high temperatures required for the activation anneal.

1.2.2 Valley Splitting

A more important consideration for implementing MOS vs. Si/SiGe quantum dot qubits is the valley splitting. In bulk silicon, electrons reside in six degenerate valleys, i.e. six symmetric minima in the conduction band in k-space. When electrons are confined to a 2D interface, like the Si/SiO$_2$ interface or within the Si quantum well in a Si/SiGe substrate, the six-fold degeneracy is lifted such that electrons preferentially reside in the two valleys perpendicular to the interface. The energy splitting between the two remaining valleys is known as the valley splitting, or $\Delta_{VS}$.

To initialize spin qubits and to perform spin dependent gate operations, the valley splitting must be large in relation to $k_B T$ and the qubit energy [52]. For quantum dot devices within a dilution refrigerator, the electron temperature is typically of order 100 mK or 8.7 µeV. For pure spin qubits, the qubit energy is given by the Zeeman splitting of the electron spin, $g\mu_B B = 116$µeV/T × B, where $\mu_B$ is the Bohr magneton
and the g-factor is around 2. In other quantum dot qubit architectures, the qubit consists of a two electron spin system such that the qubit energy is now determined by the singlet-triplet splitting instead of the single spin Zeeman splitting [26, 53]. As a rule of thumb, a sufficient valley splitting for high fidelity spin operations is about 100 μeV [52]).

In general, MOS quantum dot systems have observed larger valley splittings (∼100-1,000 μeV) [54, 55, 34, 35] than their SiGe counterparts (∼10-100 μeV) [32, 56]. In fact the low observed valley splitting in SiGe systems is a serious limitation that has motivated recent materials engineering to boost the valley splitting. Theoretical predictions of the valley splitting in ideal SiGe wells predict a high value assuming an atomically sharp interface [57]. However, atomic level disorder at the Si/SiGe interface, e.g. atomic steps and random alloy disorder, cause mixing between the valleys and severely reduce the observed valley splitting [58]. Recently, a group has demonstrated that the valley splitting in Si/SiGe quantum wells is dominated by the substrate disorder [59]. In the case of MOS, despite the more disordered Si/SiO₂ interface, the larger valley splitting is attributed to the large energy barrier of the SiO₂ which creates a sharp potential boundary condition at the interface [54].

1.2.3 Interface Disorder

One of the biggest challenge in implementing MOS quantum dots is the presence of disorder at the Si/SiO₂ interface, specifically shallow electron traps within a few meV of the conduction band edge [21, 60]. Shallow electron traps make it difficult to form well defined quantum dots and can be catastrophic for single electron manipulation, making quantum gate operations impossible [61].

Defects at the Si/SiO₂ interface have been studied for nearly half a century for classical MOS devices [49], but the defects previously studied are mostly interface
defects residing around the mid-gap of the Si band gap. In contrast, the types of defects relevant to quantum dot devices are very shallow electron traps, far away from the mid-gap, which can charge and discharge at dilution refrigerator temperatures (~100 mK). Electrons confined in mid-gap charge traps at cryogenic temperatures will be frozen-in and inert.

Furthermore, the density of shallow traps is difficult to measure with conventional measurement techniques and as such, must be measured through electron spin resonance [62].

The metric typically cited to characterize the quality of the interface is the low temperature electron mobility [13, 20, 31, 21, 25]. In MOS, typical mobilities are on the order of a few thousand cm²/Vs with exceptional samples reaching a few tens of thousands [21, 63, 36]. In this work, we have fabricated MOSFETs with peak mobilities of up to 23,000 cm²/Vs, the highest reported electron mobility for a thin-oxide (≤ 30 nm) [21]. The highest reported MOS mobility is 71,000 cm²/Vs [63]. Si/SiGe samples routinely report mobilities in the hundreds of thousands [64, 32], with the best samples reaching over a million cm²/Vs [50].

However, the peak mobility is an incomplete metric of the interface disorder relevant for quantum dot devices. This is because the electron mobility is a function of the electron density and the peak mobility occurs at relatively high densities (10¹² cm⁻² in MOS) where enough electrons are present in the system to screen out disorder and scattering centers [36]. In contrast, quantum dot devices operate in the low electron density regime (< 10¹¹ cm⁻²).

Furthermore, previous studies of shallow trap densities in high mobility MOSFETs have shown that the peak mobility does not correlate with the density of shallow traps [60]. In fact, the higher mobility sample in this study actually demonstrated a higher density of shallow traps.
In general, the nature of these shallow traps is not known. How these shallow traps are generated is also unknown. In this thesis, we demonstrate that standard annealing treatments will passivate shallow traps generated by electron-beam lithography. We will further show how measurements of the shallow trap density, which are performed by electron spin resonance, are related to other more typical metrics of the interface quality.

1.3 Outline of Thesis

The purpose of this thesis is to detail how to minimize shallow electron traps in quantum dot devices at the Si/SiO$_2$ interface and demonstrate the formation of very low disorder quantum dot devices. We demonstrate that annealing treatments are effective in passivating shallow defects created by high-energy processes and show that the shallow defect density correlates with conventional transport measurements in the low electron density regime. Ultimately we demonstrate a promising qubit architecture in MOS for controllably defining uniform quantum dots.

In the first section we detail the fabrication and characterization the highest reported mobility thin-oxide MOSFETs in order to leverage this process to fabricate low-disorder quantum dot devices. To first order we use the electron mobility as a metric of interface disorder. Because the quality of the Si/SiO$_2$ interface is critically dependent on the details of processing, we explore a myriad of various process splits and their effect on our device’s low temperature (4.2 K) electron mobility. We discover a surprising method of boosting the device’s low temperature mobility by annealing the sample in forming gas at low temperature in the presence of aluminum.

With an optimized process flow in place, we examine the effect of electron-beam lithography on the quality of the Si/SiO$_2$ interface, utilizing continuous wave (CW) electron spin resonance measurements to directly probe shallow electron traps.
Electron-beam lithography is a ubiquitous method of fabricating quantum dots in research laboratories and is known to create damage at the Si/SiO$_2$ interface. We find that a forming gas anneal is sufficient to reverse the damage created by e-beam lithography. In addition we find that our low temperature transport measurements of the devices’ percolation threshold, i.e. the lowest electron density required to support a conductive pathway, agree with our lowest temperature (360 mK) ESR measurements of the shallow trap density. This demonstrates agreement between two independent methods of assessing the oxide interface.

Finally, leveraging the above process optimizations, we fabricate and characterize a MOS quantum double dot device in a reconfigurable device architecture with an integrated charge sensor. We demonstrate the formation of uniform, lithographically defined dots that can be tuned down to the single electron regime. We measure a $1/f$ charge noise characteristic of $3.4 \mu$eV/Hz$^{1/2}$ at 1 Hz at 300 mK, commensurate with other reported charge noise values in silicon devices. We also measure a valley splitting of $110 \pm 26 \mu$eV, large enough to support high-fidelity spin selective operations.
Chapter 2

Introduction to the MOS System

2.1 The MOSFET

The primary type of device studied in the first part of this thesis are n-type inversion MOSFETs with a degenerately n-doped poly-silicon gate. The basic structure of this kind of MOSFET consists of a boron doped (p-type) crystalline silicon substrate where an insulating SiO$_2$ layer is thermally grown on the (100) surface (Figure 2.1 (a)). The oxide is capped with a metallic gate which is commonly made from aluminum or degenerately doped poly-silicon. On either side of the metallic gate, source and drain contacts are formed in the silicon substrate by degenerately doping the silicon with an n-type dopant like phosphorus or arsenic. When a voltage is applied to the gate metal, a conducting channel of electrons forms at the Si/SiO$_2$ interface.

The starting substrate used in this thesis is a commercially grown gate stack produced by Novati Technologies. The substrate is high-resistivity (1000-3000 $\Omega$-cm) float-zone (100) p-type silicon. This resistivity corresponds to a doping density of about $10^{13}$ cm$^{-3}$ of boron. 30 nm of dry, chlorinated thermal oxide was grown at the surface and then capped with 200 nm of amorphous silicon (a-Si). With these substrates, we then define the gate pattern and implant arsenic to degenerately dope
2.1 MOSFET Fabrication

the source, drain and gate to form the basic MOSFET structure. The device is annealed at high temperature to activate the dopants and repair damage incurred from the high energy implant. Finally aluminum contacts are thermally evaporated and the device receives a final forming gas (95% N\textsubscript{2}, 5% H\textsubscript{2}) anneal to improve the contact resistance and passivate interface defects.

2.2 MOSFET Operation

At low temperatures (4.2 K), with no voltage applied to the gate, the MOSFET is off, meaning conduction between the source and drain is impossible. This type of MOSFET is known as an enhancement mode MOSFET. In contrast, if the MOSFET is on with 0 applied voltage to the gate, the MOSFET is referred to as a depletion mode MOSFET.
For enhancement mode MOSFETs with 0 applied gate voltage, the Fermi level in the silicon layer resides somewhere within band-gap so that no electrons populate the conduction band (Figure 2.1 (b) (c)). By applying a positive voltage to the gate in reference to the drain, a field is applied across the oxide layer and raises the Fermi level within the silicon layer. When the Fermi level rises above the conduction band edge at the Si/SiO$_2$ interface (Figure 2.2 (b), (c)), free electrons from the n-doped source and drain are able to populate the conduction band at the Si/SiO$_2$ interface below the gate. This creates what is known as an inversion layer of electrons at the interface which enables conduction from the source to the drain. When conduction is dominated in a MOSFET by the minority carrier (electrons for a p-type substrate), the device is said to be in inversion mode.

This section will review several important MOSFET concepts including the threshold voltage, the MOSFET capacitance, and current voltage (I-V) characteristics.

### 2.2.1 Threshold Voltage

The gate voltage at which the conduction is turned on is known as the threshold voltage ($V_{th}$). Monitoring the threshold voltage is a quick and convenient way of detecting charge at the Si/SiO$_2$ interface and is an easy way to screen devices. For our Si/SiO$_2$/poly-silicon gate stacks, we consistently measure threshold voltages around 70 mV at 4.2 K. A threshold voltage that differs significantly from this value is a good indication that something went wrong during processing.

At low temperatures, the threshold voltage is simply the gate voltage at which the Fermi level crosses the silicon conduction band edge at the Si/SiO$_2$ interface. The threshold voltage can be calculated from the work functions, i.e. the difference between the material’s equilibrium Fermi energy and the vacuum level, of the gate
metal ($\phi_m$) and the silicon substrate ($\phi_{Si}$):

$$V_{th} = E_g + (\phi_m - \phi_{Si}) - Q_{ox}/C_{ox}$$  \hspace{1cm} (2.1)

Here $E_g$ is the silicon bandgap in volts, 1.1 V, $Q_{ox}$ is charge present at the Si/SiO$_2$ interface, and $C_{ox}$ is the oxide capacitance. $Q_{ox}$ can have several origins and will add a voltage offset to the calculated threshold voltage. We will review charges in the oxide in detail in section 2.4.

For p-type silicon at cryogenic temperatures, the Fermi level resides $\sim$10 meV above the valence band. For a degenerately n-doped poly-silicon gate metal, the Fermi level resides $\sim$30 meV below the conduction band edge within an impurity band edge created by the n-dopants. This yields a threshold voltage on the order of tens of mV which is in good agreement with our measured threshold voltage of 70 mV. This is an indication that there is little charge present at the Si/SiO$_2$ interface.

At room temperature, conduction turns on well before the Fermi level crosses the conduction band edge because a significant number of electrons can be thermally excited into the conduction band edge. In addition, the equilibrium Fermi level within p-type silicon tends to be closer to the midgap, increasing with higher temperature. At room temperature, the threshold voltage can be calculated by:

$$V_{th} = \frac{eN_aW_{max}}{C_{ox}} + (\phi_m - \phi_{si}) + 2\phi_{fp} - \frac{Q_{ox}}{C_{ox}}, \hspace{1cm} (2.2)$$

$$\phi_{fp} = \frac{k_B T}{e} \ln \left( \frac{N_a}{n_i} \right) \hspace{1cm} (2.3)$$

$$W_{max} = \left( \frac{4\epsilon_{Si} \phi_{fp}}{eN_a} \right)^{1/2} \hspace{1cm} (2.4)$$

In expressions 2.2-2.4, $N_a$ is the silicon doping density, $\phi_{fp}$ is difference between the bulk silicon Fermi level and the silicon mid-gap, $n_i$ is the intrinsic carrier concentration $\approx 1.5 \times 10^{10}$ cm$^{-3}$, $\epsilon_{Si}$ is the relative permittivity in silicon 11.9$\epsilon_0$, and $W_{max}$ is the
maximum width of the depletion region. The depletion region width will be discussed in further detail in the next section.

2.2.2 Capacitance

Measuring the MOSFET’s capacitance as a function of gate voltage is a powerful way of extracting many useful parameters of the device like the substrate doping density and type, electron density, threshold voltage, and oxide thickness. In addition, it can be used to measure charge at the Si/SiO₂ interface and calculate interface state defect densities.

When the MOSFET is in inversion, the device acts simply as a parallel plate capacitor where the metal gate and the inversion layer act as the two plates. The
MOSFET capacitance is given simply by the oxide capacitance $C_{ox}$:

$$C_{ox} = \frac{\epsilon_{ox}}{d}, \quad \epsilon_{ox} = 3.9\epsilon_0$$  \hspace{1cm} (2.5)$$

Here $d$ is the oxide thickness. For a 30 nm oxide, the oxide capacitance is $C_{ox}=115$ nF/cm$^2$. It is often useful to think of the oxide capacitance in terms of electron density per volt which yields $C_{ox}/e=7.18 \times 10^{11}$ cm$^{-2}$V$^{-1}$.

With the oxide capacitance we can estimate the electron density $n$ as a function of gate voltage:

$$n = C_{ox} \times (V_G - V_{th})$$  \hspace{1cm} (2.6)$$

It should be noted however that at low temperatures, the electron density is not actually 0 at $V_{th}$ due to disorder [65, 36], thus this estimate of the electron density will underestimate the actual electron density. Depending on the sample, the electron density at the threshold voltage is around $10^{11}$ cm$^{-2}$ due to disorder at the interface. This will be discussed in detail in chapter 4.4.2.

For gate voltages below $V_{th}$, the MOSFET capacitance has more complicated behavior (Figure 2.3). For gate voltages far below $V_{th}$, where the Fermi level at the Si/SiO$_2$ interface lies below the valence band, holes accumulate at the interface in the valence band. This regime is known as accumulation. Changes to the gate voltage in this regime simply add or remove holes from the interface. Therefore, the MOSFET capacitance is again the oxide capacitance.

Once the gate voltage is raised high enough such that the Fermi level no longer resides below the valence band, holes depopulate the valence band at the Si/SiO$_2$ interface and a condition occurs known as the flatband voltage. This is the gate voltage at which the potential is constant throughout the silicon substrate, and as such the energy bands are “flat” spatially. The flatband voltage $V_{fb}$ is simply the
Figure 2.3: Simplified picture of the MOSFET capacitance as a function of gate voltage. Three regimes of accumulation, depletion and inversion are shown [66].

difference in the metal and silicon work functions:

\[ eV_{fb} = \phi_m - \phi_{Si} \quad (2.7) \]

When the gate voltage is raised higher than the flatband voltage, dopants in the bulk substrate begin ionizing, starting with acceptors near the Si/SiO\(_2\) interface. The region of ionized dopants in the bulk silicon is known as the depletion region. As the gate voltage increases further, the width of the depletion region increases.

A simple model for the depletion region width \( W_{dep} \) can be expressed in terms of the acceptor doping density \( N_a \) and the surface potential \( \phi_s \), i.e. the difference in the Fermi level in the bulk silicon and the Fermi level at the Si/SiO\(_2\) interface:

\[ W_{dep} = \sqrt{\frac{2\varepsilon Si \phi_s}{e N_a}} \quad (2.8) \]
The approximation made here is that within the depletion region width, all acceptors are ionized and outside the depletion region, all acceptors are neutral.

The capacitance of the depletion region $C_{\text{dep}}$ is therefore:

$$C_{\text{dep}} = \frac{\varepsilon_{\text{Si}}}{W_{\text{dep}}} = \frac{\varepsilon_{\text{Si}}}{\sqrt{\frac{2\varepsilon_{\text{Si}}\phi_s}{eN_a}}}$$  \hspace{1cm} (2.9)

The total capacitance $C$ of the MOSFET in depletion is then:

$$\frac{1}{C} = \frac{1}{C_{\text{ox}}} + \frac{1}{C_{\text{dep}}}$$  \hspace{1cm} (2.10)

$$C = \left[ \frac{1}{C_{\text{ox}}^2} + \frac{2(V_G - V_{fb})}{eN_a\varepsilon_{\text{Si}}} \right]^{-1/2}$$  \hspace{1cm} (2.11)

Here, $(V_G - V_{fb})$ has been substituted for the surface potential $\phi_s$.

As the gate voltage is raised further, eventually the MOSFET enters inversion, as the gate voltage is raised above the threshold voltage and the MOSFET capacitance is once again the oxide capacitance.

Figure 2.4 is a measured CV curve of a MOS capacitor fabricated from our gate stack. In inversion and accumulation, we see that the capacitance saturates to the oxide capacitance. We also see that the depletion approximation is consistent with our experimental data in the depletion regime but the error is substantial around the flatband voltage. The full analytic solution for the MOS capacitance involves solving Poisson’s equation for the surface potential while taking into account the thermal spread of the depletion region boundary.

Fortunately, a convenient expression for the flatband capacitance can be found by linearizing Poisson’s equation around the flatband voltage. The flatband capacitance $C_{fb}$ is thus given by:

$$C_{fb} = \frac{1}{\left(\frac{d}{\varepsilon_{\text{ox}}} + \frac{L_D}{\varepsilon_{\text{Si}}} \right)} ,$$  \hspace{1cm} (2.12)
Figure 2.4: Typical measured quasi-static CV curve for a 1.4 mm$^2$ MOS capacitor. $V_{fb}$ is around -1V and $V_{th}$ is around -0.3V.

\[ L_D = \sqrt{\frac{k_B T \epsilon_{si}}{e e N_a}} \]  

(2.13)

$L_D$ in expression (2.12) is the Debye length which characterizes the screening length of the electric field by electrons within the semiconductor.

Expression (2.12) is useful to experimentally determine the flatband voltage which is useful in extracting information of defects at the interface. Deviations of the experimentally measured flatband voltage from the expected flatband voltage are an indication of defects and charges in the oxide. We will see in section 2.4 that the CV measurement is a very powerful tool in measuring classical oxide defects. Unfortunately, it is insufficient to detect shallow electron traps.
Figure 2.5: Drain current measured at room temperature for a Si/SiO$_2$/Al$_2$O$_3$/Al MOSFET for $V_G = 0 - 1$ V in 0.1 V intervals. Also indicated is the saturation source drain voltage $V_{SD}^{sat}$.

### 2.2.3 Current-Voltage (I-V) Characteristics

Above threshold, conduction through the MOSFET channel is dictated by both the source-drain voltage $V_{SD} = V_S - V_D$ and the gate voltage $V_G$. For a constant gate voltage above threshold $V_G > V_{th}$, as $V_{SD}$ is increases from 0, the drain current $I_D$ increases. As $V_{SD}$ increases further, the voltage drop at the drain contact to the gate decreases until the voltage drop between the drain and gate is reduced below the threshold voltage. At this point, the drain current saturates with increasing $V_{SD}$. The saturation voltage is defined as $V_{SD}^{sat} = V_G - V_{th}$. So in effect, as $V_{SD}$ increases, the MOSFET channel at the drain terminal pinches off.
The ideal current-voltage relationship for $I_D$ is given by:

\[
\text{Non-saturation: } I_D = \frac{W\mu C_{ox}}{2L} \left[2(V_G - V_{th})V_{SD} - V_{SD}^2\right] \tag{2.14}
\]

\[
\text{Saturation: } I_D = \frac{W\mu C_{ox}}{2L} (V_G - V_{th})^2 \tag{2.15}
\]

Here, $W$ and $L$ are the MOSFET channel width and length, respectively. $\mu$ is the effective electron mobility. The above expressions make the assumption that the electron mobility is constant, the channel is rectangular, the oxide charge is constant, the channel is strongly inverted, and that the lateral electric field is much less than the vertical field. We should note that at low temperatures, the electron mobility assumption breaks down as the mobility changes as a strong function of the electron density.

Figure 2.5 illustrates the room temperature family of drain current curves for a Si/SiO$_2$/Al$_2$O$_3$/Al MOSFET. The threshold voltage for this device is about 0.3 V which is higher than our typical Si/SiO$_2$/poly-silicon gate stack due to the presence of fixed negative charge at the SiO$_2$/Al$_2$O$_3$ interface.

### 2.3 The 2 Dimensional Electron Gas

To a good approximation, the electrons residing in the inversion layer at the 2D Si/SiO$_2$ interface are non-interacting and as such are often referred to as a 2 dimensional electron gas, or 2DEG. These electrons are free to move in the plane of the interface, but are confined in the perpendicular direction by a triangular potential well formed by the bending of the conduction band edge and the barrier of the SiO$_2$ (Figure 2.2 (c)). In the perpendicular direction, the confinement creates quantized energy sub-bands where the first excited state is of order 10 meV above the ground state. At liquid helium temperatures, electrons only reside in the first sub-band. A
A convenient analytic expression for the ground state wavefunction is given by [36]:

$$
\psi_0 = \sqrt{\frac{1}{2}} \left( \frac{3}{z_0} \right)^{\frac{3}{2}} z e^{-\frac{3z}{2z_0}}
$$

(2.16)

Here $z_0$ is the average penetration of the wavefunction into the silicon and is found variationally by minimizing the ground state energy.

### 2.3.1 Density of States

The density of states for a 2D electron system is given by:

$$
D(E) = 2g_v \frac{1}{(2\pi)^2} 2\pi k \frac{dk}{dE}
$$

(2.17)

Here, a spin degeneracy of 2 is added and $g_v$ is the valley degeneracy which is 2 for the two occupied valleys perpendicular to the interface. These valleys are preferentially populated because their effective mass ($m_l = 0.92m_0$) in the direction of confinement is larger than the other four valleys'. Recalling that the orbital energy $E_0$ is proportional to $\frac{1}{m^*}$, the perpendicular valleys’ ground state orbital energy is thus lowered by at least ten meV below the four parallel valleys.

For the typical parabolic energy dispersion, where $E_0$ is the energy of the lowest sub-band, the electron energy $E$ is given by:

$$
E = E_0 + \frac{\hbar k^2}{2m_l}
$$

(2.18)

The density of states is constant with electron density and is given by:

$$
D(E) = \frac{2m_l}{\pi\hbar^2}, \quad E > E_0
$$

(2.19)

$$
= 0, \quad E < E_0
$$

(2.20)
This yields a value of $D(E) = 1.6 \times 10^{11}$ cm$^{-2}$meV$^{-1}$.

### 2.4 SiO$_2$ Defects

In general, there are four accepted types of oxide charges that have been extensively studied in the MOS literature for classical MOS devices: mobile ionic charge ($Q_M$), oxide trapped charge ($Q_{OT}$), fixed oxide charge ($Q_F$), and interface trapped charge ($Q_{IT}$ or $D_{IT}$) \[67, 49\]. In this section we will briefly review these types of defects, their origins, methods of measuring them, and methods of reducing them. We will then introduce shallow electron traps, i.e. the defect most relevant for quantum MOS devices but least studied of all the other defects. We will see that standard methods of measuring classical defects like CV measurements are insufficient to measure shallow electron traps, motivating electron spin resonance measurement techniques.

#### 2.4.1 Mobile Ionic Charge

**Characteristics**

Mobile ionic charges in the oxide occur because of contamination by alkali metal ions like Na$^+$, K$^+$, and Li$^+$. Na$^+$ is the most common. As their name suggests, these ions can freely diffuse through the silicon dioxide layer at room temperature. The mobility of these charges increases with decreasing ionic radius.

In classical devices, mobile ionic charge creates instabilities in the threshold voltage. A negative threshold voltage shift occurs the closer the charges are to the Si/SiO$_2$ interface. For a 2D concentration $N_M$ of mobile ionic charges, the threshold voltage can shift up to $\Delta V_{th} = eN_M/C_{ox} = N_md_{ox}(\mu m)/2.15 \times 10^{10}$. For our 30 nm oxide, a concentration of $7 \times 10^9$ cm$^{-2}$ will result in a 10 mV shift in threshold voltage.

Over time, the mobile ionic charge distribution in the oxide will change as the ions diffuse. Furthermore, the application of an electric field from the gate voltage

will cause the ions to drift from the Si/SiO$_2$ interface to the SiO$_2$/gate interface and vice versa. Thus, the normal operation of the MOSFET will cause hysteretic shifts to the threshold voltage.

In addition to creating threshold voltage shifts, the presence of mobile ionic charge in the oxide will affect the transport properties of the 2DEG. Mobile ionic charges will act as Coulomb scattering centers which can significantly lower the electron mobility. At low electron densities, ionic charges near the Si/SiO$_2$ interface will act as electron traps with a binding energy of 15-30 meV [36]. At low temperatures, mobile ionic charges may contribute to parasitic disorder dots in quantum dot devices.
Figure 2.7: Schematic of threshold voltage shift in the CV curve. This shift can be induced by drifting mobile ionic contaminants through the oxide with an applied field or by injecting charge into bulk oxide traps.

**Measurement Methods**

A CV measurement scanned from low gate voltage to high gate voltage will reveal the presence of mobile ionic charges through a hysteric threshold voltage shift (Figure 2.7). A high gate voltage at the end of the scan will drift mobile ions from the SiO$_2$/gate interface to the Si/SiO$_2$ interface. A subsequent CV scan from high to low gate voltage will demonstrate a parallel negative shift of the CV curve. Often times this measurement is done while heating the sample to 100-200 C to increase the mobile ionic drift velocity. This method of biasing and heating the sample is known as a bias temperature stress (BTS) measurement.

A similar method as the above CV BTS measurement will also reveal differences in the low temperature electron mobility. The sample is similarly heated to 100-200 C while biasing the gate negatively to pull mobile ions to the gate. The sample is
then immediately cooled down in a cryostat with the gate voltage biased negatively and the electron mobility is measured. In the presence of mobile ionic contaminants, this method will boost the electron mobility by a significant percentage.

Minimization Methods

Contamination from mobile ionic charges can be minimized by careful processing of the samples. Contamination in wet processes (from dirty glassware or sodium or potassium containing chemicals) can easily transfer to an exposed silicon oxide. In addition, contaminated furnace tubes or boats can contaminate samples during high temperature processes. RCA cleaning is an effective way of cleaning mobile ionic contaminants from the surface of a sample and is absolutely necessary to avoid transferring contamination to the furnace tube (which, if contaminated, will subsequently contaminate future samples). An ambient chlorine gas like dichloroethane (DCE) or dichloroethylene can be flowed through a contaminated furnace tube to clean it. In addition, DCE can be introduced during the oxide growth process itself to help combat mobile ionic contamination. Chlorine is incorporated into the oxide directly and can bind to and neutralize alkali ions.

2.4.2 Oxide Trapped Charge

Characteristics

Oxide trapped charge are electron or hole traps that exist in the bulk of the oxide, often near the Si/SiO\textsubscript{2} interface. These traps are electrically insulated from the silicon substrate as well as the gate. Because these traps are for the most part electrically inaccessible, they have no associated capacitance. Charge can, however, be injected into these traps from the silicon substrate through large voltage pulses or by optically induced hot electrons. These traps are not important when the device
is operated at cryogenic temperatures and in the absence of large voltage pulses. In any case, they can be generating by ionizing radiation.

**Measurement Methods**

A CV measurement can be performed before and after injecting charge into oxide traps by electron avalanche injection or hot-electron injection. The effect of charging the oxide traps is to shift the CV curve along the voltage axis (Figure 2.7). The difference threshold voltage before and after charge injection will yield an equivalent oxide trapped charge density $\Delta N_{OT} = \frac{C_{ox} \Delta V_{th}}{e}$.

**Minimization Methods**

Oxide trapped charge can be minimized by reducing the sample’s exposure to ionizing radiation whenever possible, for example by depositing metals via thermal evaporation instead of electron-beam evaporation. Trapped charge from ionizing radiation can be neutralized or compensated generally by low temperature annealing ($\sim 400$ C) in forming gas, however high temperature annealing ($\sim 900$ C) is required to remove trap sites [49, 67].

**2.4.3 Fixed Oxide Charge**

**Characteristics**

Fixed oxide charge is positive charge located approximately 2.5 nm within the Si/SiO$_2$ interface [69, 67]. It is associated with structural defects at the Si/SiO$_2$ interface and its magnitude is dependent on the silicon orientation as well as the oxidation conditions. It is a stable defect in the sense that the magnitude of charge is fixed at room temperature and, like oxide trapped charge, is not a function of the silicon surface potential. Fixed oxide charge will shift the threshold voltage by $-Q_F/C_{ox}$. 
Measurement Methods

The magnitude of $Q_F$ can be extracted from the flatband voltage $V_{fb}$ measured from a CV sweep:

$$Q_F = (\phi_{metal} - \phi_{Si} - V_{fb})C_{ox} \quad (2.21)$$

The above expression is simply the difference in the ideal flatband voltage ($\phi_{metal} - \phi_{Si}$) and the measured flatband voltage. Notice that in this measurement, the flatband voltage, instead of the threshold voltage, is used as a reference for the voltage shift caused by the fixed charge. This is because the presence of interface trapped charge will also shift the threshold voltage away from the ideal threshold voltage but will not shift the flatband voltage. At the flatband voltage, no interface states are charged.

Minimization Methods

The oxidation growth temperature and subsequent annealing conditions can reduce the magnitude of fixed oxide charge, following the so called “Deal-Grove Triangle” \[69, 49\]. Figure 2.8 illustrates the dependence of fixed oxide charge on the oxidation temperature and annealing temperature. For higher oxidation temperatures, the fixed charge density decreases (hypotenuse). At a given oxidation temperature, the fixed oxide charge can be reduced by annealing the sample at the oxidation temperature in an inert ambient gas like nitrogen or argon (vertical edge). Cooling the wafer in an ambient gas will not change the fixed charge density.

The orientation of the silicon surface on which the oxide is grown is also important to minimizing the fixed oxide density. The (100) silicon surface yields the lowest fixed oxide charge density because the number of bonds through this crystal plane at the Si/SiO$_2$ interface is minimized. The (111) surface yields the most fixed oxide charge...
Figure 2.8: Schematic of the Deal-Grove triangle for (111) silicon. (100) silicon $Q_f$ values are about 1/3 of the measured values for (111). Higher dry oxidation temperatures result in lower fixed charge densities (diagonal line), referred to here as “Surface-State Charge.” Subsequent annealing in argon or nitrogen can reduce fixed charge densities (vertical lines). (from B.E. Deal, “characteristics of the surface-state charge (Qss) of thermally oxidized silicon” J. Electrochem. Soc. Vol 114, No. 3 (March 1967) pp. 266-274)

density. For this reason, most device grade silicon oxides are grown on the (100) silicon surface.

2.4.4 Interface Trapped Charge

Characteristics

Interface trapped charge, or interface states, are defect states that exist within the silicon bandgap at the Si/SiO$_2$ interface $^{70}$ $^{49}$ $^{67}$. They may be positively (donor-like) or negatively (acceptor-like) charged and are electrically accessible to free charges
in the silicon. Thus as the surface potential increases, more interface states become charged.

Interface states are associated with broken symmetry and structural defects at the Si/SiO₂ interface. These may consist of dangling bonds or other defect complexes at the interface. The density of interface states have been shown to be correlated with the density of $P_b$ centers [71], a well known Si/SiO₂ defect consisting of a silicon atom back bonded to three other silicon atoms with a dangling bond. Like fixed oxide charge, the density of interface states is a function of the silicon surface orientation with the (100) surface yielding the lowest interface state density. Interface states can be created through ionizing radiation or similar dangling-bond creating processes [67, 49]. In addition, metallic contaminants can create interface states.

The interface state density is typically represented by $D_{IT}$ which has units of areal density per energy ($\text{cm}^{-2}\text{eV}^{-1}$). Because interface states can exist throughout the bandgap, it is important to account for the depth of the trap within the bandgap.

**Measurement Methods**

The standard way of measuring interface state densities is to measure a MOS capacitor’s low frequency (quasi-static) CV measurement and compare the resulting capacitance curve to a high frequency CV curve [67].

The low frequency CV curve will reflect the charging of interface states as the number of filled or empty states changes as the Fermi level is swept across the bandgap. This is represented by an additional series capacitance at the Si/SiO₂ interface which “smears” out (or “stretches-out”) the low frequency capacitance curve relative to the ideal curve (Figure 2.9).

Interface states will only contribute to the device’s low frequency, or “quasi-static” (QS), capacitance if the measurement frequency is slower than $\tau_{IT}^{-1}$, where $\tau_{IT}$ is the interface trap time constant. In other words, the interface states must have enough
Figure 2.9: Schematic of the high frequency (HF) MOS CV curve compared to the low frequency (LF) CV curve. Interface states add to the measured device capacitance in the low frequency measurement causing the CV curve to smear out. The high frequency CV curve is not sensitive to the charging of interface states. At inversion and above, the high frequency capacitance saturates at the minimum capacitance value.

time to charge during the measurement in order for the interface state to contribute to the measured capacitance. $\tau_{IT}$ is the timescale for a thermally excited electron to be captured by an interface state which is calculated by the following expression:

$$
\tau_{IT} = \frac{1}{\bar{V}\sigma_n n_i} e^{\exp \left[ \frac{e(\phi_B - \phi_s)}{k_B T} \right]}
$$

Here, $\bar{V}$ is the mean thermal electron velocity, $\sigma_n$ is the interface trap cross section $\approx 6 \times 10^{-16} \text{cm}^{-2}$, $n_i$ is the intrinsic carrier concentration, and $\phi_B$ is the bulk potential.

The QS capacitance is then compared to a high frequency ($\sim$1 MHz) capacitance measurement. Now the measurement frequency is much higher than $\tau_{IT}^{-1}$ so that the interface states do not have time to respond to the AC modulation. From the thresh-
old voltage and above, the high frequency capacitance saturates at the minimum of the capacitance dip where the depletion region width is maximized (Figure 2.9). In this regime, the only charges fast enough to respond to the AC modulation are majority carriers (holes) at the edge of the depletion region which charge and discharge acceptors. Minority carriers (electrons) at the inversion layer cannot respond fast enough to the modulation to add to the measured capacitance. This measurement must be done with MOS capacitors (no source or drain) because it relies on the bulk silicon being an inefficient supplier of electrons.

With the high and low frequency capacitances \( C_{HF} \) and \( C_{LF} \), the interface state density \( D_{IT} \) as a function of the surface potential \( \phi_s \) is given by:

\[
D_{IT}(\phi_s) = \frac{C_{ox}}{e} \left( \frac{C_{LF}}{C_{ox} - C_{LF}} - \frac{C_{HF}}{C_{ox} - C_{HF}} \right) \text{ cm}^{-2}\text{eV}^{-1} \tag{2.23}
\]

The surface potential can be calculated as a function of the gate voltage by the following integral:

\[
\phi_s(V_G) = \Delta + \int_{V_{fb}}^{V_G} dV_G \left[ 1 - \frac{C_{LF}}{C_{ox}} \right] \tag{2.24}
\]

In expression 2.24, \( \Delta \) is a constant which is obtained by calculating the voltage offset of the measured flatband voltage in \( C_{LF} \) to the ideal flatband voltage.

The above expressions will give valid \( D_{IT} \) values in the regime below the threshold voltage and above the surface potential value at which \( \tau_{IT} \) equals the AC modulation period. For a 500 kHz modulation, we can measure interface states \( \sim 250 \text{ meV} \) away from the valence band edge. It is important to note that it is very difficult to probe interface states very close \( \leq 100 \text{ meV} \) to the band edges. Low temperatures are required for the surface potential to be near the band edge at the onset of inversion which exponentially increases the trap timescale. Electron spin resonance is an effective way of probing interface states shallower than this.
Minimization Methods

Interface states can be effectively annealed out in a low temperature (400 C) forming gas anneal \[49, 72, 70, 73, 67\]. Hydrogen from the forming gas diffuses through the gate stack and passivates dangling bonds at the Si/SiO\(_2\) interface. 400 C is the optimal annealing temperature in terms of interface state minimization as two competing mechanism come into play during the anneal. With higher annealing temperatures, hydrogen can diffuse more quickly through the gate stack. However, at the same time, dangling bonds that have been passivated by hydrogen during the anneal can thermally de-passivate the bonded hydrogen with higher temperature \[67\]. After a forming gas anneal, \(D_{IT}\) can reach densities in the low \(10^{-10} \text{cm}^{-2} \text{eV}^{-1}\).

2.4.5 Shallow Electron Traps

The defect most relevant for quantum dot devices are shallow electron traps \[60, 21\]. These traps can be thought of as interface states that reside a few meV below the conduction band edge, whereas typical interface states are more than 100 meV away from the conduction band edge. Few studies have been performed on shallow traps as they are too close to the band edge to be measured through CV measurements.

Shallow traps are electrically active at cryogenic temperatures, meaning they can charge and discharge during the operation of a quantum dot device. These traps behave as parasitic quantum dots and interfere with forming well defined electrostatic quantum dots. Deeper interface traps are frozen in and only contribute a static electric field.

Little is known about the exact nature of shallow traps. They may be associated with band tail states from broken symmetry at the Si/SiO\(_2\) interface \[74\] or they may also originate from individual positive charges a few nanometers within the oxide \[20\]. A priori, it is unknown whether or not standard annealing procedures are effective in reducing the shallow trap density, or how shallow trap densities relate to other
measurements of the oxide interface. We will explore these topics in the following chapters.
Chapter 3

Engineering a Low-Disorder Si/SiO$_2$ Interface in MOSFETs

In this section we study the effect of fabrication processes on the quality of the Si/SiO$_2$ interface in silicon MOS Field-Effect Transistors (MOSFETs). As mentioned before, the Si/SiO$_2$ interface is susceptible to defect formation by high-energy processes during device fabrication, so we explore various processing parameters in order to minimize Si/SiO$_2$ interface disorder in order to leverage these processing parameters later for quantum dot devices. We fabricate MOSFETs in Hall bar geometries and measure the low temperature (4.2 K) electron mobility. Hall bars can be fabricated with high device yield (as the device can be made arbitrarily small to reduce gate leakage) and the electron mobility can be measured quickly.

The low temperature Hall mobility $\mu$ is a commonly used but incomplete measure of the Si/SiO$_2$ interface [13]. It is related to the conductivity $\sigma$ of the system by:

$$\mu = \frac{\sigma}{ne}$$  \hspace{1cm} (3.1)
where $n$ is the 2D electron density. It can also be expressed in terms of the mean electron scattering time $\bar{\tau}$ such that

$$\mu = \frac{e\bar{\tau}}{m^*}$$

$$\bar{\tau}^{-1} = \tau_{Coulomb}^{-1} + \tau_{SR}^{-1} + \tau_{Phonon}^{-1} + \ldots$$

Expression 3.3 is the sum of various scattering mechanisms (Coulomb, surface roughness, phonon...) and is known as Matthiessen’s rule [36].

At higher temperatures the electron scattering and thus electron mobility is dominated by phonon scattering. However, below temperatures of about 10 K, phonons are strongly suppressed and electron scattering is dominated by Coulomb scattering centers (i.e. charges at the Si/SiO$_2$ interface) and surface roughness. As a function of electron density, the low temperature mobility first increases as more electrons fill the system and are able to screen out Coulomb scattering centers [36, 75, 76]. However, as the electron density is increased by applying a larger gate voltage, the increased electric field pulls the 2DEG tighter to the interface and surface roughness scattering begins to dominate. Consequently, at higher electron densities the mobility decreases. This gives rise to the characteristic mobility peak shape that is observed at liquid helium temperatures (4.2 K).

Because the mobility is measure of electron scattering and thus interface charges and defects, it may be tempting to use the low temperature peak mobility as a metric for interface quality. However, there are many factors at play that are folded into the electron mobility that may obscure the factors relevant for high-quality quantum dot devices [76, 77]. For one, the electron mobility occurs at high electron densities which is an unrealistic scenario for quantum dot devices operating in the single-electron regime. Additionally, in a previous study we have observed in two separate high
mobility devices that the device that demonstrated the higher peak mobility also demonstrated a higher density of shallow traps [60].

For now we will proceed to use the electron mobility as a proxy for the interface disorder with the ansatz that high electron mobility is probably necessary but not sufficient for a low density of shallow traps. Once the process is optimized in terms of electron mobility, we study large-area MOSFETs (which have significantly lower device yield than Hall bars) in Chapter 4 in order to measure the shallow trap density with electron spin resonance. We will also correlate the shallow trap density with the minimum conducting electron density (or percolation threshold density), measured at 300 mK.

In this section we will detail an optimized fabrication process which consistently yields low temperature electron mobilities in the range of 16,000-17,000 and as high as 23,000 cm$^2$/Vs. We find that the peak electron mobility is a strong function of various fabrication parameters and discover that annealing in the presence of aluminum yields a surprisingly large boost in the mobility.

3.1 Devices Fabricated

The devices we studied in this section are inversion MOSFETs fabricated in a Hall bar geometry (Figure 3.1). Hall bars are simply MOSFETs fabricated in a double cross geometry with six ohmic contacts. The contacts are arranged to be able to measure voltage drops in the 2DEG parallel and perpendicular to the direction of current. Our samples are typically 200 µm wide with a 3.8 mm intercontact spacing. Typically a known current $I_{xx}$ is pushed through the end terminals (1-4) of the device with $V_G > V_{th}$. Then the parallel voltage drop $V_{xx}$ is measured across the terminals 2-3 or 6-5 to measure the conductivity of the 2DEG as a function of gate voltage. A magnetic field is then turned on and the electron density is then measured by
Figure 3.1: Schematic of the Hall bar geometry with an optical microscope image of a fabricated Hall bar device mounted to an 8 pin DIP header.

measuring the perpendicular voltage drop $V_{xy}$, i.e. the Hall voltage, across terminals 2-6 or 3-5. The electron density is also measured as a function of the gate voltage and this measurement can be used to back out the oxide capacitance. The typical voltages that are measured are in the range of 10-100 µV so the measurements are done with a lock-in amplifier. From the measured values of the conductivity and electron density, we can calculate the electron mobility.

The base fabrication process for these devices follows three guidelines to minimize defect formation at the Si/SiO$_2$ interface and maximize the electron mobility:

1. Minimize high-energy processes during device fabrication.
2. Eliminate sources of mobile ionic contamination.
3. Anneal to passivate damage incurred during processing.

Regarding the first guideline, many fabrication processes generate high-energy photons and electrons which can create interface states [78, 79, 49, 80]. These processes include but are not limited to: reactive ion etching, sputtering, electron-beam lithography, electron-beam evaporation, ion implantation. For our process, we will
utilize only low power reactive ion etches (<50 W) and evaporate metals only via thermal evaporations.

To minimize sources of mobile ionic contamination, we try to keep the oxide encapsulated as much as possible. This motivated our decision to start with a commercially grown oxide stack with 200 nm a-Si gate deposited by the foundry. In addition, we only use fused quartz glassware to clean the sample prior to being annealed at high temperature. Normal borosilicate glassware contains a high percentage of sodium (≈5% by volume) which has the possibility of leeching out of the glassware and onto the sample [68]. At high temperatures, contaminants on the surface of the sample or within the furnace tube can diffuse into the oxide. Finally, we have a dedicated thermal evaporator for the evaporation of 99.9999% (6N) pure aluminum only for making contacts to the source, drain, and gate of devices. Dirty metal sources can be a source of metallic and ionic contamination [81].

Despite minimizing the exposure to high-energy processes, a few high energy processes are still required during device fabrication, notably ion implantation to dope the contacts and gate, and reactive ion etching to define the gate geometry. After these processes we perform anneals to passivate damage incurred [82, 83].

Starting from our commercially grown gate stack, our base process is as follows (Figure 3.2):

1. Reactive ion etch (RIE) and buffered oxide etch to define contact holes

2. Blanket arsenic ion implantation (35 keV, 5×10^{15} \text{ cm}^{-2}) to degenerately dope contacts and gate (Leonard Koko Inc.)

3. Reactive ion etch to define gate geometry

4. High temperature (900 C, 1 hour, N_2) anneal to activate the implant and repair ion implantation damage in PRISM Tystar oxidation furnace

5. Deposit aluminum on contacts and gate
6. Forming gas anneal (435 C, 25 minutes, 5%H₂) to improve contacts and reduce interface states in PRISM Thermco metals anneal furnace or CVD 1034 furnace (tube 2)

7. Deposit gold solder pads

Individual devices are then diced and screened at room temperature for gate leakage. Devices are made in batches of four and usually two or three of the four demonstrate acceptable levels of gate leakage (<1 µA at 1 V). Good devices are then mounted to an 8 pin DIP header connecter with a dab of vacuum grease and gold wires are hand soldered to the contact pads of the device.

### 3.2 Experimental Setup

#### 3.2.1 Cryostat

The devices are mounted to the insert of an Oxford continuous flow helium cryostat sitting within a Varian electromagnet (Figure 3.3). The cryostat pulls cold helium gas continuously from a liquid helium Dewar, achieving temperatures slightly below 4.2
K in continuous mode. In single shot mode, the cryostat can achieve temperatures below 2 K for about 30 minutes. The cryostat is filled with liquid helium and then sealed off from the liquid helium dewar. This volume of liquid helium is then pumped on by a mechanical pump, cooling the system to below 2 K. The cooldown time for the flow cryostat is about 30 minutes and consumes $\leq 1$ liter of liquid helium per hour, allowing for quick turn around of device measurement.

The cryostat insert is mounted on a goniometer which can rotate the sample in-situ. The sample is oriented perpendicular to the magnetic field in order to measure the electron density via the Hall voltage.

### 3.2.2 Instrumentation and Methods

With the sample cold, a semiconductor parameter analyzer is used to check the low temperature device characteristics. The gate leakage for a good device is typically on the order of 10’s of picoamps or less and the threshold voltage at 4.2 K is typically 60-70 mV. Each contact is checked to ensure its resistance is ohmic at low temperature. Additionally, the saturation current is recorded as a function of the gate voltage.
The experimental setup for measuring the mobility consists of a Princeton Applied Research (PAR) lock-in amplifier with a differential voltage pre-amplifier, a Stanford Research Systems (SRS) SIM928 isolated voltage source, an HP34401A digital voltmeter and a Varian electro-magnet (Figure 3.3). Each instrument is interfaced to a common PC via GPIB or RS-232 to USB and controlled via scripts written in Matlab.

The reference channel voltage output ($V_{\text{mod}}$) of the lock-in amplifier is used in conjunction with a 9.5 MΩ resistor ($R_0$) to supply current source of $I_{xx} \approx 100 \text{ nA}$ at a low frequency between 3 Hz and 17 Hz. The lock-in frequency is usually chosen to be some prime number in order to avoid picking up harmonics (especially 60 Hz). The SIM928 voltage source is used to scan the gate voltage and the two inputs of the differential pre-amplifier are attached the device contacts of interest. The output of the lock-in is digitized through the HP34401A voltmeter and fed to the PC.

The above resistor and modulation voltage values are chosen such that $R_0$ is much larger than the MOSFET channel resistance ($R$) and $I_{xx} = V_{\text{mod}}/R_0$ is much less than the MOSFET saturation current in the measurement regime. This presents a trade-off of being able to measure the device properties at very low electron densities (where the channel is very resistive) vs. signal strength and measurement time. 100 nA with the above values turns out to be a good compromise at 4.2 K and allows for fast measurements down to electron densities of about $2 \times 10^{11} \text{ cm}^{-2}$. To probe the device’s transport properties at lower electron densities, we simply use a larger resistor.

$V_{xx}$ is measured at 0 field as a function of the gate voltage to measure the device conductivity $\sigma$. Usually it is a good idea to measure $V_{xx}$ on leads 2-3 and compare those values to $V_{xx}$ measured on leads 5-6. The two should give identical values to within a few percent. If the values from 5-6 differ greatly from 2-3, it may be a sign that the device is very asymmetric or the interface quality is very low yielding large
inhomogeneities in the 2DEG. The conductivity is calculated simply by:

\[ \sigma(n) = \frac{I_{xx} L}{V_{xx} W} \]  

(3.4)

Here, \( L \) is the distance between leads 2 and 3 (3.8 mm) and \( W \) is the channel width (200 \( \mu \)m).

\( V_{xy} \) is then measured in a 0.5 T magnetic field (\( B_0 \)) as function of the gate voltage. \( V_{xy} \) should be the same measured on 2-6 and 3-5. Additionally, \( V_{xy} \) should go to 0 at 0 magnetic field, but sometimes a small residual voltage can be measured at 0 magnetic field. This is usually attributed to a small component of \( V_{xx} \) being measured due to slight device asymmetries. In this case, this offset can simply be subtracted out. The sample can also be rotated by 90 degrees with the goniometer to calibrate the sample orientation in the field. The goniometer is rotated until the 2DEG plane is parallel to the magnetic field and \( V_{xy} \) is measured to be 0. Rotating the sample 90 degrees from this point will then orient the 2DEG exactly perpendicular to the field. The electron density \( n \) is calculated by:

\[ n = \frac{1}{10,000} \frac{I_{xx} B_0}{e V_{xy}} \text{ cm}^{-2} \]  

(3.5)

An easy way of checking that the measurement is set up correctly initially is to measure \( V_{xy} \) at 0.5 T with \( V_G \) set 1.0 V above \( V_{th} \). With a 30 nm oxide, one would expect an electron density of about \( 7 \times 10^{11} \text{ cm}^{-2} \) which should yield a \( V_{xy} \) value of 58 \( \mu \)V.

Figure 3.4 is representative data taken of a typical device’s electron density as a function of gate voltage. The data is linearly fit to extract the device’s capacitance which is found to be \( 6.71 \times 10^{11} \text{ cm}^{-2} \text{V}^{-1} = 107 \text{ nF cm}^{-2} \) which is in good agreement with the nominal capacitance of a 30 nm oxide, 115 nF cm\(^{-2}\).
Figure 3.4: Measurement of electron density from $V_{xy}$ as a function of the gate voltage.

At low gate voltage, the electron density appears to slightly deviate away from the linear fit. In this case this effect is due to the MOSFET channel resistance ($R$) becoming of similar order to the 9.5 MΩ resistor ($R_0$) on the output of the AC modulation. In this regime, the actual current flowing through the device is no longer constant as a function of $V_G$ and is divided down by a factor of $\frac{1}{1+R/R_0}$. The measured $V_{xy}$ is subsequently divided down by $\frac{1}{1+R/R_0}$ which yields a fictitiously high electron density calculated by expression $3.5$. A similar effect will occur if the modulation voltage is set large enough to where $I_{xx}$ is larger than the saturation current of the channel. In later experiments where we probe the low electron density regime, we instead use an 867 MΩ resistor.
3.3 Results

In this section we summarize low temperature (4.2 K) mobility data of devices produced by our base process yielding mobilities of upwards of 16,000 cm$^2$/Vs. We examine some key process splits to examine the magnitude of reactive ion etching (RIE) damage and additionally explore the effect of annealing the devices in the presence of aluminum (so called “alnealing”) \[70, 84\].

We compare the effect of etching the polysilicon layer to define the gate geometry via RIE and compare those mobility results to a device etched via wet chemical etching. Wet chemical etching produces no high-energy photons or electrons, unlike RIE, so comparing these two process splits elucidates the effect of the RIE damage on the Si/SiO$_2$ interface. In order to observe the full extent of the RIE damage, we perform the RIE at the very end of the fabrication process with no subsequent annealing, as annealing is known to repair RIE damage \[85, 86, 87\]. We find that the RIE minimally lowers the electron mobility and that performing the RIE of the polysilicon layer at the very end of the device fabrication process actually makes the device less susceptible to sodium contamination.

Finally we explore a surprising trick that significantly boosts the electron mobility: a low temperature forming gas anneal with aluminum deposited on top of the polysilicon gate. Aluminum is a commonly used material for gate metals and is known to lower interface state defect densities by cracking H$_2$ into atomic H which is extremely efficient at passivating interface defects \[70\]. However, aluminum can also contain trace amounts of alkali ionic contaminants which can diffuse into the gate oxide if deposited directly onto the oxide \[81\]. By utilizing a polysilicon gate (deposited in a clean foundry) and depositing aluminum on top of the polysilicon, we find the devices benefit from the annealing action of the aluminum without being contaminated by alkali ionic contaminants.
3.3.1 Base Process Mobility

Figure 3.5 illustrates a typical mobility curve of a sample processed with our base process measured at 4.5 K. This sample demonstrates a peak mobility of 16,900 cm$^2$/Vs at an electron density of $9.8 \times 10^{11}$ cm$^{-2}$. The mobility curve reflects the expected peak shape where at low electron densities the mobility rapidly increases before decreasing at higher electron densities. At electron densities lower than $9 \times 10^{11}$ cm$^{-2}$ the mobility is limited by Coulomb scattering centers which are screened out as the electron density increases. At around $9 \times 10^{11}$ cm$^{-2}$ the mobility levels off and at higher densities (necessarily corresponding to higher electric fields) the mobility decreases due to surface roughness scattering.
3.3.2 Damage from Reactive Ion Etching

We now examine the damage effects caused by reactive ion etching. Reactive ion etching is necessary to pattern nano-scale features in quantum dot devices, as nano-scale features patterned by wet chemical etching are dominated by the crystal grain boundaries leading to rough, uncontrolled etch profiles.

Within our base fabrication process, we perform two anneals (a high temperature N$_2$ anneal and a low temperature forming gas anneal) after the reactive ion etch of the polysilicon layer to repair the damage incurred by the etch. For the following experiments, we modify our fabrication process so that the polysilicon RIE occurs at the very end of the process, after the high temperature and forming gas anneals. In this manner we will be able to measure the un-annealed damage effects of the RIE on a device ("Reactive ion etch damage"). We further compare these mobility results to a device where the polysilicon layer was etched by a wet chemical etch.

Figure 3.6 plots the mobility curves of these three samples as function of electron density at liquid helium temperatures. Of these three devices, the reactive ion etch damaged sample demonstrates the lowest mobility with a peak mobility of 14,900 cm$^2$/Vs. We note that this mobility is still very good. The next highest mobility sample is the base process, again with a peak mobility of 16,900 cm$^2$/Vs. The highest mobility sample is the wet chemical etched device with a peak mobility of 19,200 cm$^2$/Vs. This is consistent with what we expect. The RIE generates defects at the interface which lowers the peak mobility slightly. The base process sample demonstrates a slightly higher peak mobility where the RIE-generated defects have been annealed. The highest mobility device is the wet chemical etched device where defects have not been generated at all by an RIE.

As a general trend, we observe that as the peak mobility increases, the peak density $n_{\text{peak}}$ (the electron density corresponding to the peak mobility) decreases. In addition, we observe that at high electron densities, the mobility of different devices
Figure 3.6: Comparison of 4.2 K mobilities for various polysilicon etch parameters.

converge. This suggests that at high electron densities, these devices demonstrate similar levels of surface roughness. The difference in peak mobility originates then from differing amounts of Coulomb scattering centers that are generated during the fabrication process. Therefore, in general, lower peak densities are an indication of fewer defects in our samples. The device characteristics of these three process splits are tabulated in Table 3.1.

Plotted at the bottom in Figure 3.6 is another sample fabricated with the base process that was contaminated by sodium during the high temperature anneal. This anneal was performed in the PRISM Tystar oxidation tube prior to the tube being cleaned with Dichloroethylene. The peak mobility of the sodium contaminated sample is 6,800 cm²/Vs at a density of 13.2×10¹¹ cm⁻². This device’s peak mobility is more
Table 3.1: Mobility comparison of polysilicon etch parameters at 4.2 K

<table>
<thead>
<tr>
<th>Process Split</th>
<th>Peak Mobility (cm²/Vs)</th>
<th>Peak Density (10¹¹ cm⁻²)</th>
<th>% Deviation in Peak Mobility from Base</th>
</tr>
</thead>
<tbody>
<tr>
<td>Base process</td>
<td>16,900</td>
<td>9.8</td>
<td>–</td>
</tr>
<tr>
<td>Wet chemical etch</td>
<td>19,200</td>
<td>6.7</td>
<td>+13.6</td>
</tr>
<tr>
<td>Reactive ion etch damage</td>
<td>14,900</td>
<td>9.2</td>
<td>-11.8</td>
</tr>
<tr>
<td>Base process with sodium contamination</td>
<td>6,800</td>
<td>13.2</td>
<td>-59.7</td>
</tr>
</tbody>
</table>

than a factor of two lower than the base process device and the peak density is significantly higher.

We confirm the presence of sodium in this sample by performing a bias temperature stress and re-measuring the low temperature mobility. A -2 V bias is applied to the gate while the sample is heated at 200 C on a hotplate for 10 minutes to drift Na⁺ ions from the Si/SiO₂ interface to the SiO₂/gate interface. The sample is then quickly transferred into the cryostat the electron density and mobility are re-measured.

Figure 3.7 plots the mobility data for the sodium contaminated sample before and after the bias temperature stress. After the BTS, the peak mobility is boosted significantly from 6,800 to 10,000 cm²/Vs and the peak density is lowered from 13.2 to 10.9×10¹¹ cm⁻². A boost in mobility and a reduction in the peak density are both indications that charge has been reduced at the Si/SiO₂ from the BTS, consistent with drifting mobile ionic contamination from the Si/SiO₂ interface to the gate. The fact that the mobility does not fully recover to the base process mobility may be an indication residual mobile ionic contaminants remain at or have diffused back to the Si/SiO₂ interface.

To quantify the amount of sodium that has been drifted away from the oxide interface, we can examine the electron density data as a function of gate voltage.
as shown in Figure 3.8. After the BTS, the electron density curve is shifted in the positive $V_G$ direction. We can extrapolate the gate voltage where the electron density is 0 ($V_G^0$) for both devices. From the shift in $V_G^0$ of both devices we can thus estimate the concentration of sodium in the oxide, assuming all of the sodium prior to the BTS was at the oxide interface and after the BTS was at the gate. This yields a mobile ionic contaminant concentration of $N_M = C_{ox} \Delta V_G^0/e = 6.2 \times 10^{10}$ cm$^{-2}$. The transport parameters of the pre and post BTS sodium contaminated sample are tabulated in Table 3.2.

Table 3.2: Mobility comparison of a sodium contaminated sample before and after a bias temperature stress

<table>
<thead>
<tr>
<th>Process Split</th>
<th>Peak Mobility (cm$^2$/Vs)</th>
<th>Peak Density ($10^{11}$ cm$^{-2}$)</th>
<th>$V_G^0$ (V)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre BTS</td>
<td>6,800</td>
<td>13.2</td>
<td>-0.057</td>
</tr>
<tr>
<td>Post BTS</td>
<td>10,000</td>
<td>10.9</td>
<td>+0.036</td>
</tr>
</tbody>
</table>

It should be noted that several batches of devices were fabricated with the RIE etch damage process flow (i.e. with the oxide encapsulated in polysilicon until the last step) while the source of sodium contamination was being investigated. Each of these devices achieved peak mobilities consistently in the 14,000-16,000 cm$^2$/Vs range. Meanwhile, the devices fabricated with the base process (i.e. the oxide surrounding the polysilicon gate was exposed prior to the high temperature anneal) were contaminated and achieved mobilities in the 6,000-8,000 cm$^2$/Vs range. After the furnace was cleaned with dichloroethylene (known commercially as Trans-LC), the base process achieved mobilities in the range of 17,000 cm$^2$/Vs.

What was likely happening was that sodium contamination present in the furnace was diffusing into the surface of the exposed oxide of base process devices and then was diffusing laterally through the oxide into the channel region under the polysilicon gate (see Figure 3.1). In the devices where the gate was not etched prior to the high
Figure 3.7: Comparison of mobilities of a sodium contaminated sample pre and post bias temperature stress.

temperature anneal (the RIE etch damage process), the polysilicon encapsulated the oxide and acted as an effective diffusion barrier to the sodium present in the furnace. So while performing the RIE at the end of the process generates some un-annealed defects, it makes the process much more robust to sodium contamination. This provides a reliable way of fabricating high mobility samples which, in the context of processing, provides a useful tool to track down sources of sodium.

3.3.3 The Effect of Aluminum Annealing

The final set of process splits we will examine demonstrate a surprisingly large boost in the low temperature mobility when the samples are annealed with aluminum covering
Figure 3.8: Comparison of electron density as a function of $V_G$ of a sodium contaminated sample pre and post bias temperature stress.

$\approx 90\%$ of the degenerately doped polysilicon gate. A layer of aluminum on top of the relatively resistive polysilicon gate is required when performing ESR measurements on large-area MOSFETs. This aluminum layer acts as a microwave shunt and prevents microwaves from being absorbed by the polysilicon gate \[88\]. For device and process consistency, we also evaporated aluminum on top of the polysilicon gate of our Hall bars (Figure 3.1). This section will detail how we deduced that the presence of aluminum during a forming gas anneal dramatically boosts the peak mobility.

The devices studied in this section were fabricated such that the etch of the polysilicon layer to define the gate geometry was performed last. As described in the previous section, a small amount of defects are generated when the polysilicon is etched via RIE in the final step, but overall this makes the process much more
robust to sodium contamination which can be much more detrimental to the devices. A set of devices was then fabricated in this manner with aluminum evaporated onto the gate (referred to as “Aluminum over gate, RIE”). A second set was fabricated identically with no aluminum evaporated onto the gate (referred to as “No Aluminum over gate, RIE”). Both devices were then subsequently annealed in forming gas and etched via RIE.

Figure 3.9 shows the difference in mobility curves of these two devices. The device with the aluminum is the same device as the “RIE etch damage” device from the previous section and again demonstrates a peak mobility of 14,900 cm²/Vs. The device fabricated without the aluminum yields a maximum measured mobility more than 10 times lower, 1,000 cm²/Vs at an electron density of $26 \times 10^{11}$ cm⁻². The mobility for this device was still very slowly increasing at the maximum measured electron density, so the actual peak mobility is at an even higher density than what was measured. Based on the slow rate of increase in mobility as a function density though, the value of the peak mobility is probably not much higher than 1,000. These device characteristics are listed in Table 3.3.

It was not immediately obvious why we observed such a large difference in mobility by the simple fact that aluminum was present on the gate for the last several steps of the fabrication process. We postulated several explanations for the reduction in mobility in the absence of aluminum:

1. The aluminum on the polysilicon gate was shielding the underlying oxide from radiation produced by the RIE.

2. The degenerately doped polysilicon was actually highly resistive and therefore applying an inhomogeneous field over the area of the channel.

3. The aluminum was assisting in the annealing process as has been observed in extremely high efficiency silicon photovoltaic devices and early MOS devices.
Through the following process splits, we determined that the aluminum is likely assisting in low temperature forming gas anneal.

First we explore whether or not the presence of aluminum on top of the polysilicon is shielding the underlying oxide from radiation produced from the RIE during the polysilicon etch. Each previous high-mobility device either had its polysilicon layer etched while aluminum covered the polysilicon gate region or experienced a high temperature anneal after the polysilicon etch. Perhaps the absence of aluminum over the gate was allowing radiation to penetrate through the polysilicon gate into the underlying oxide and since the polysilicon was being etched in the final step, many defects were being generated that were not subsequently annealed. To test this, we fabricated another Hall bar without aluminum covering the gate and performed a wet
chemical etch of the gate region. These results are also plotted in Figure 3.9, labelled “No aluminum over gate, wet chemical etch.” The only difference between this device and the “No aluminum over gate, RIE” is the method of etching.

Table 3.3: Mobility comparison of aluminum annealed Hall bars

<table>
<thead>
<tr>
<th>Process Split</th>
<th>Peak Mobility (cm²/Vs)</th>
<th>Peak Density (10¹¹ cm⁻²)</th>
<th>% Deviation in Peak Mobility from Base</th>
</tr>
</thead>
<tbody>
<tr>
<td>Annealed with Al over gate, RIE</td>
<td>14,900</td>
<td>9.2</td>
<td>-11.8</td>
</tr>
<tr>
<td>Annealed without Al over gate, RIE</td>
<td>1,000*</td>
<td>26*</td>
<td>-94.0</td>
</tr>
<tr>
<td>Annealed without Al over gate, wet chemical etch</td>
<td>2,900</td>
<td>30</td>
<td>-82.8</td>
</tr>
</tbody>
</table>

*Maximum measured values

The wet chemical etched device in this case demonstrates a peak mobility of 2,900 cm²/Vs at 30×10¹¹ cm⁻² (listed in table 3.3). The mobility of this device is slightly higher than the RIE etched device without aluminum, but nowhere near the mobility of the base process device. Therefore, we rule this explanation out.

The second explanation suggests the n+ degenerately doped polysilicon maybe very resistive. A very resistive gate material could be trapping charge, creating an inhomogeneous 2DEG by applying an inhomogeneous electric field. To test this, we etched the aluminum off of a device fabricated via the base process and remeasured its low temperature mobility. None of its transport characteristics changed. Additionally, the resistivity of the gate was measured, yielding ∼0.005 Ω-cm which is definitively conductive.

We therefore conclude that most likely the aluminum is assisting in the annealing of Si/SiO₂ interface defects in forming gas. This effect has been studied extensively in the context of interface state defects of MOS capacitors [70, 82] as well as in the
context of the surface recombination velocity of silicon solar cells [84], but never before in the context of the low temperature mobility.

The likely mechanism is that aluminum, a highly reactive metal, cracks molecular H\textsubscript{2} present in the forming gas into atomic H [70]. Atomic H is much more reactive than H\textsubscript{2} and is able to efficiently bond to dangling bonds at the Si/SiO\textsubscript{2} interface. This annealing procedure with aluminum is referred to in the literature as “alnealing” [70, 82] In the study of MOS capacitors, the capacitors fabricated with aluminum gates demonstrated much lower interface state defect densities compared with other gate materials. Furthermore, depositing a layer of silicon nitride beneath the aluminum gate in MOS capacitors negated the benefit of annealing the sample with aluminum. Silicon nitride is one of the few materials that hydrogen will not diffuse through. In the photovoltaic studies, alnealing was required to achieve efficiencies of 19%. We should note that low interface state densities have been achieved with polysilicon gates [73] although the annealing time to achieve the minimum interface state density is on the order of 1 hour instead of 10’s of seconds with an aluminum gate [72].

3.4 Conclusion

In this chapter we have detailed how to consistently fabricate high mobility MOSFETs to leverage this process to fabricate low defect density quantum dot devices. In the next chapters we will measure the shallow trap density of high mobility MOSFETs and correlate our transport measurements with our electron spin resonance measurements.

Within our process we find that our reactive ion etch does minimal damage to the Si/SiO\textsubscript{2} interface and find that sodium contamination can potentially be much more damaging to the quality of our devices. To minimize sodium contamination, it is critical to run an HCl clean of the high temperature annealing furnace prior to annealing the samples. Performing the RIE of the gate geometry as the last step
of the process keeps the oxide encapsulated during all other fabrication steps and is useful to minimize sodium contamination at the cost of incurring slight RIE damage. Finally we show that a large boost in the low temperature mobility can be achieved by annealing the sample in forming gas with a polysilicon/aluminum gate metal stack.
Chapter 4

Annealing Shallow Electron Traps

In the previous chapter we have developed a consistent process yielding high mobility MOSFETs. As the previous chapter detailed, the quality of the Si/SiO$_2$ interface depends critically on the fabrication parameters of the device. Moving forwards we would like to leverage this process in order to fabricate low disorder MOS quantum dot devices which will require the use of electron-beam (e-beam) lithography. E-beam lithography is known to create interface state defects at the Si/SiO$_2$ interface which can be annealed out through high and low temperature annealing treatments. However, it is not known what the effect of the e-beam is on the generation of shallow traps, or whether or not the standard annealing treatments are effective.

This chapter will detail several experiments published in reference [21] examining the effect of e-beam lithography on the shallow trap density at the Si/SiO$_2$ interface and the efficacy of annealing treatments of the e-beam generated shallow traps. We fabricate high mobility MOSFETs (23,000 cm$^2$/Vs) for electron spin resonance (ESR) measurements and Hall bars for transport measurements and subject them to a typical e-beam lithography dose (10 keV, 40 μC/cm$^2$) and a subsequent forming gas anneal. We directly measure the shallow trap density of our large area MOSFETs using ESR and compare these results to the low temperature mobility measurements.
of our Hall bars. We will introduce the basics of the ESR measurement of shallow traps and will introduce the percolation threshold density, a measurement of the lowest electron density required to support a conductive pathway which can be extracted from low temperature transport measurements. With these tools we show that the ESR measured shallow trap density at the lowest measured temperature is consistent with the percolation threshold density, demonstrating agreement between a specialized measurement (ESR) and a conventional one (transport). We further demonstrate that a standard forming gas anneal is sufficient to passivate shallow defects generated by the e-beam. The devices studied here demonstrate the highest reported mobility (23,000 cm$^2$/Vs) for a thin-oxide ($\leq$30 nm) MOSFET.

### 4.1 Introduction

A critical processing step for quantum dot devices that we have neglected up to now is electron-beam (e-beam) lithography. E-beam lithography is the "workhorse" of nano-fabrication in research labs, but is known to generate interface defects in MOS devices. Quantum dot devices rely on nano-scale electrostatic gates in order to provide tight spatial confinement so that quantum effects can be observed at experimentally achievable temperatures. Dots in silicon must be made especially small in order to observe quantum effects due to their large effective mass, compared to GaAs. Dots in MOS have typical feature sizes of order 50 nm.

As an aside, industrial foundries, such as Intel, utilize immersion lithography systems which achieve nanometer resolution by immersing samples in a liquid with a refractive index greater than one in order to enhance the diffractive limit. Patterning with these systems is extremely complicated as the photomasks used to pattern samples must precisely engineer the waveform of the exposure light in order to beat the diffraction limit. Additionally, multiple exposures and patterns are required for
each layer of lithography. Immersion lithography systems are prohibitively expensive for academic and research laboratories, selling in the range of $30-50 million, not to mention the cost and complications in writing and producing masks used in the systems themselves.

An affordable alternative method of fabricating nanoscale devices is nanoimprint lithography [89], which does not incur the radiation damage of e-beam lithography. However, alignment of multiple layers of nanoimprint lithography, while possible [90], is non-trivial and most modern quantum dot designs rely on several layers of nanolithography. Therefore, the workhorse tool for nanoscale patterning in research and academic laboratories is e-beam lithography.

E-beam tools operate by firing a focused beam of high energy (10-100 keV typically) electrons at an e-beam sensitive polymer resist. The energetic electrons cause the long polymer chains in the e-beam resist to break down and become soluble in developing solution. The de Broglie wavelength of electrons at these energies is of order 0.01 nm so the diffraction limit of an e-beam is much lower than that of an optical system. However, the true resolution of an e-beam system is usually limited by 1) scattering of the electron beam within the substrate and 2) space charge effects of the beam spot itself. With some effort, one can achieve feature sizes less than 10 nm.

The same energetic electrons that are used to pattern the nanoscale gates in a quantum dot device are also known to create interface defects which can trap electrons near the quantum dot and make single electron operations impossible. Figure 4.1 illustrates a SRIM calculation of 10 keV electron beam penetrating our device gate stack with visualizations of variations in the conduction band edge leading to shallow traps. Shallow electron traps are the defect most damaging to the operation of quantum dot devices. These traps reside only a few meV below the conduction band edge and can be thought of as spatial variations in the conduction band edge.
Figure 4.1: Left: simulation of a 10 keV electron beam scattering through our device stack. Center: disorder landscape cartoon of the conduction band edge at the Si/SiO₂ interface after a typical e-beam dose. The energy scale of the defects are on the order of \( \leq 10 \) meV. Right: diagram of the MOS band diagram indicating the location of shallow electron traps.

or as interface states residing very near the conduction band edge. They are only apparent at cryogenic temperatures as at warmer temperatures these traps become indistinguishable from the conduction band itself. At helium temperatures, shallow traps are electrically active, meaning they can trap and de-trap conduction band electrons. In contrast, traditional mid-gap interface states are not electrically active as electrons trapped in these states are frozen in. The energy scale of shallow traps coincides with the typical charging energy of a quantum dot. As the quantum dot plunger voltage is increased to increase the electron population of the dot, nearby shallow traps are also populated. These are also referred to in the literature as defect dots or parasitic dots. Nearby trapped electrons can exchange with the QD electron and can cause decoherence and interfere with Pauli spin blockade [61].

Many studies have been performed on the creation of interface states by e-beam irradiation in MOS capacitors but most of these studies are performed well above liquid helium temperatures and probe mid-gap (instead of shallow) electron traps [91, 92, 93, 94]. There have been only a few studies regarding annealing on the performance in MOS quantum dot devices [20, 95]. These studies qualitatively demonstrate
that a forming gas anneal after e-beam lithography significantly improves the performance of the quantum dot compared to an un-annealed QD \cite{20, 95}. However, it is difficult to glean statistics of shallow defect generation and annealing from mesoscopic measurements.

By performing a bulk ESR measurement of large-area MOSFETs that have been irradiated with an e-beam and annealed, we directly measure the efficacy of annealing treatments on generated shallow traps. In addition to performing ESR measurements on large-area MOSFETs, we also measure the transport characteristics of Hall bars that have been similarly irradiated and annealed. Up until now we have used the low temperature peak mobility as a proxy for the interface quality, knowing that it is an incomplete metric of the Si/SiO$_2$ interface quality. It would be useful to understand how transport measurements relate to measurements of the shallow trap density, keeping in mind that higher peak mobility does not necessarily correlate with lower shallow trap density. We find that transport measurements of the percolation threshold (extracted from transport measurements) correlate with the shallow trap density and will show data confirming that a forming gas anneal is effective in passivating shallow traps generated by the e-beam exposure.

### 4.2 Devices Fabricated

The devices measured in this experiment are n-channel inversion MOSFETs following the base fabrication process developed in the previous chapter. Two sets of devices (which will be referred to as samples A and B) were fabricated where each set of devices consists of a Hall bar for transport measurements and a large-area MOSFET ($\sim 1$ cm$^2$) for ESR measurements (Figure 4.2). One set will be irradiated by an e-beam and subsequently annealed (set B) and the other will serve as the control set and will remain un-irradiated (set A). In this section we will first detail the fabrication
processes of the two sets of devices fabricated, then discuss the details of our e-beam
dose. Finally we will comment on the design of our large-area ESR MOSFETs.

Both sets of devices were fabricated following the base process developed in the
previous chapter up until the high temperature anneal. After the high temperature
anneal the process was split and one set of devices (samples B) received a blanket
e-beam exposure (10 keV, 40 $\mu$C/cm$^2$) while the other set (samples A) served as the
control set and remained unexposed. This process split is shown in Figure 4.3. The
e-beamed samples were spin coated with 300 nm of ZEP520a e-beam resist prior to
irradiation to ensure proper alignment of the exposure. The calculated penetration
depth of 10 keV electrons through our resist and device stack is $\approx$ 1 $\mu$m (Casino[96]),
deep enough to damage the Si/SiO$_2$ interface. After stripping the e-beam resist,
both sets of devices were metallized with 300 nm of thermally evaporated Al over the
source/drain and n$^+$ poly-Si gate. Both sets of devices were then annealed in forming
gas (5% H$_2$, 435°C) for 25 minutes. Finally, Ti/Au was thermally evaporated onto the
contacts for soldering to the device. In addition to the two sets of devices mentioned
above, a third Hall bar (sample C) was fabricated to demonstrate the damage created
by the e-beam exposure. This sample was fabricated identically to sample A but was
Figure 4.3: Diagram of process splits performed for this experiment. Device A is the result of the control process. Device B was exposed to an e-beam dose and was subsequently annealed in forming gas. Device C was exposed to the e-beam dose as the final step of processing with no post-exposure anneal.

then coated with e-beam resist and received an e-beam exposure (identical to sample B) at the very end of processing, with no post-exposure anneal.

Our choice of e-beam energy (10 keV) and dose (40 µC/cm²) was chosen based on the clearing dose and total exposure time of the e-beam resist used (ZEP520a). Typically the clearing dose goes down as a function of beam energy as more electrons are able to backscatter which assist in the exposure. At low (∼10 keV) e-beam energy the exposure is primarily achieve through backscattered electrons whereas at high e-beam energy (∼ 100 keV) the exposure is achieved through forward scattered...
electrons, thus a larger (i.e. longer) dose is necessary for higher energy beams. On the other hand, much smaller feature sizes are achievable with higher energy beams.

For a 10 keV beam energy, the resist clearing dose is 40 µC/cm². Using a 3.5 nA beam current, the total exposure time for four large area MOSFETs (where the typical device yield is roughly one in four) is about 14 hours which is experimentally achievable. For a 125 keV beam, the resist clearing dose is 360 µC/cm². This would mean the total exposure time would be about 126 hours which is experimentally unfeasible. One may argue that our exposure conditions are not comparable to a realistic quantum dot fabrication process where a 100 keV beam is often used. We argue that the relevant energy scale here is around 10 eV (the SiO₂ bandgap) because defect formation relies on the creation of electron-hole pairs within the oxide. Therefore in terms of defect creation, a 10 keV beam is probably not fundamentally different from a 100 keV beam.

It is not obvious which case would generate more total defects. In the low energy beam case, backscattered electrons allow more energy to be deposited into the oxide (creating interface defects) whereas in the high energy beam case, the majority of the energy is deposited into the bulk silicon. However, in the high energy case, the electrons that do get absorbed into the oxide may generate more defects compared to low energy electrons.

The design of our large-area MOSFETs was motivated by considerations of our ESR measurement. Our measurements are performed by a commercial ESR spectrometer with a sensitivity of $\sim 10^8 - 10^9$ unpaired spins. We can relate the number of unpaired spins ($n_\uparrow - n_\downarrow$) induced by a magnetic field to the 2D density of states $D(E)$ and the Zeeman splitting $\Delta_Z$ by

$$n_\uparrow - n_\downarrow = \frac{D(E)\Delta_Z}{2}, \quad (4.1)$$

$$\Delta_Z = g\mu_0B \quad (4.2)$$
Here \( g \) is the g-factor of conduction band electrons which is 1.9999, \( \mu_0 \) is the Bohr magneton and \( B \) is \( \approx 3,400 \), corresponding to the resonance of our microwave resonator, 9.6 GHz. With these values, the Zeeman splitting is \( \Delta_Z \approx 0.04 \) meV yielding \( n_\uparrow - n_\downarrow \approx 10^9 \text{ cm}^{-2} \). Therefore our device channel area must be \( \approx 1 \) cm\(^2\) to detect our 2DEG signal.

### 4.3 Experimental Setup

#### 4.3.1 Cryostats

The experiments described in this chapter were first tested in a Pope Scientific glass cryostat with a base temperature of 1.8 K before being cooled down in a Janis \(^3\)He cryostat with a base temperature of 300 mK. The glass cryostat is a helium bath cryostat insulated by three concentric insulation jackets: vacuum, liquid nitrogen, then vacuum. Liquid helium is transferred into the inner-most helium chamber which will maintain the cryostat at 4.2 K for about 8 hours. The chamber can also be pumped down to vacuum which will cool the cryostat down to around 1.8 K for about 5 hours.

Once devices are screened, they are mounted within the sample chamber of the \(^3\)He cryostat and cooled down. The large-area ESR MOSFETs are mounted within a volume resonator in the sample chamber and the Hall bar is mounted to a cold finger fixed to the \(^3\)He pot (Figure 4.4).

The key cooling elements of the \(^3\)He system consist of the charcoal sorb, 1 K pot and the \(^3\)He pot. These three components are contained in a vacuum canister called the inner vacuum can (IVC) which is immersed in a \(^4\)He bath. The basic cooling mechanism of the \(^3\)He cryostat is the evaporative cooling of liquid \(^3\)He. \(^3\)He gas is first condensed into the \(^3\)He pot by the 1 K pot, then the charcoal sorb pumps on the liquid \(^3\)He, cooling the system to 300 mK. Once all of the liquid \(^3\)He has evaporated
and has been absorbed into the charcoal sorb, it can be released by heating the sorb to 35 K at which point the $^3$He gas can be re-condensed into liquid. By utilizing a charcoal sorb to pump on the $^3$He gas instead of an external pump, the $^3$He gas remains in a closed cycle and can be reused indefinitely.

When the system is cooled to base temperature for the first time, we introduce a small amount of $^3$He exchange gas into the sample chamber. The exchange gas is necessary to thermalize our devices to base temperature as the devices do not make direct contact with the $^3$He pot. This small amount of exchange gas is lost to atmosphere when the system is warmed up and the sample chamber is opened.

Starting from room temperature, the cryostat takes about two days to load, pump down, and cool to base temperature. Under ideal conditions, it can be kept at base
temperature for up to 24 hours before warming up to 4.2 K, at which point the $^3$He must be re-condensed. Liquid $^4$He can be continuously added to the $^4$He bath every several days to keep the cryostat running. On average, the cryostat will consume $\sim$10 liters of liquid helium per day.

### 4.3.2 Continuous Wave Electron Spin Resonance

We measure our large-area MOSFETs using a commercial continuous wave (CW) ESR spectrometer (Bruker Elexsys E 580) \cite{97} to detect the absorption signal of electron spins confined in shallow traps. A basic schematic of our ESR setup is shown in Figure 4.5 (a) showing the microwave bridge, microwave cavity and sample, and the magnet. The sample is placed in a microwave resonator which is located inside a magnet. The magnetic field $B_0$ induces a Zeeman splitting $\Delta_Z = g\mu B_0$ where microwaves resonant to the Zeeman transition are absorbed and this absorption signal is measured by the spectrometer.

Different spin species are identified by their g-factor. The absorption signal of interest for us are conduction band electrons which have a g-factor of 1.9999. This signal is also gate voltage dependent and orientation dependent due to the Rashba field \cite{98}.

Experimentally, the microwave frequency is set by the resonator frequency. The resonator serves to concentrate the microwaves, which increases the microwave coupling to the sample, and create a standing wave mode which orients the oscillating microwave magnetic field $B_1$ perpendicular to the $B_0$ field. With a fixed microwave frequency ($\sim$9.6 GHz), the magnetic field is then swept to achieve the resonant condition.

For increased sensitivity, the spectrometer additionally employs a lock-in amplifier for phase sensitive detection. As the $B_0$ magnetic field is swept, a small ($\sim$1 G) AC (100 kHz) magnetic field is added via modulation coils mounted to the resonator.
Figure 4.5: (a) Diagram of the main components of a Bruker Elexsys E 500 ESR spectrometer showing the microwave bridge, cavity, and magnet. (b) Diagram of the microwave bridge showing the microwave source, reference and signal microwave channels, cavity, and detector diode [97].

and the resulting modulated microwave absorption signal is fed to a Schottky barrier diode. The detector diode converts the incoming microwave signal into an electrical signal which is fed into a lock-in detector. The Schottky barrier diode is most sensitive at a diode current of 200 µA, so some of the microwave power is tapped off into a reference channel and supplied directly to the diode to bias the current to 200 µA. This configuration is shown schematically in Figure 4.5 (b).

### 4.4 Results

#### 4.4.1 Hall Mobility

Transport measurements were done on all three sets of devices using standard low frequency lock-in techniques. Samples A (control) and B (irradiated and annealed) were measured in a $^3$He cryostat at temperatures between 335 mK and 4.5 K using
a constant excitation current of 1.5 nA. The threshold voltage ($V_{th}$) of these two devices was measured to be $\approx 0.07$ V at 4.2 K and increases slightly with decreasing temperature to $\approx 0.2$ V at 335 mK. Sample C (irradiated and un-annealed) was measured at 4.2 K with an excitation current of 115 nA and its threshold voltage was measured to be 1.2 V, indicating approximately $10^{12}$ cm$^{-2}$ net oxide charges (interface states and fixed oxide charge) created by the e-beam exposure.

For each sample, electron densities ($n$) were calibrated by measuring the Hall resistivity in a 0.5 T field and the mobility was extracted by standard four-terminal lock-in measurements of the sample resistivity. Mobility data as a function of temperature are shown for samples A and B (Figure 4.6) and peak mobility curves for all three samples is compared in Figure 4.7.

For samples A and B, the peak mobility increases significantly from 4 K to 300 mK as the peak density decreases. This increase in the low temperature mobility below 4 K is typically only observed in high quality samples \[63\]. At lower temperatures, electron screening of Coulomb scattering centers becomes more efficient which causes an increase in electron mobility \[76\]. Sample A’s peak mobility increases from 18,000 cm$^2$/Vs at $9.4 \times 10^{11}$ cm$^{-2}$ to 23,000 cm$^2$/Vs at $6.3 \times 10^{11}$ cm$^{-2}$. Sample A
Figure 4.7: Experimental mobility ($\mu$) plotted as a function of electron density ($n$) for three studied devices. Sample A remained unexposed to the e-beam. Sample B was exposed and annealed in forming gas, recovering more than half of sample A’s peak mobility. Sample C (inset) was exposed with no post-exposure anneal and shows significant degradation to its peak $\mu$ and a high threshold $n$. Shubnikov-de Haas oscillations are visible in samples A and B.

Demonstrates the highest reported electron mobility for a MOSFET with an oxide thickness of 30 nm or thinner \[36, 13\]. Similarly, sample B’s peak mobility increases from 12,000 cm$^2$/Vs at $8.3 \times 10^{11}$ cm$^{-2}$ to 14,000 cm$^2$/Vs at $6.1 \times 10^{11}$ cm$^{-2}$. At the lowest temperatures measured, both samples A and B demonstrate Shubnikov-de Haas oscillations consistent with the electron density.

The lowest temperature mobility data measured for samples A, B and C are compared in Figure 4.7. Sample C demonstrates the lowest mobility of $< 1,000$ cm$^2$/Vs at $n = 5.5 \times 10^{12}$ cm$^{-2}$, illustrating the damage incurred by the e-beam. This sample was only measured at 4 K and not at 300 mK as we are confident that the
mobility would not increase enough to be comparable to samples A or B. Comparing the mobility data for all three devices show that the e-beam dose significantly degrades the oxide interface (sample C) and that a forming gas anneal is sufficient to restore an e-beamed device to high mobility (sample B).

### 4.4.2 Percolation Threshold

Peak mobility, however, is measured at relatively high electron densities where the two dimensional electron gas (2DEG) can effectively screen out scattering centers \[36\] and as such is not necessarily a useful indicator of the oxide interface quality for quantum devices operating in the few electron regime \[60\]. An alternative method used to assess the interface quality from transport measurements is to fit the measured conductivity \((\sigma)\) to a percolation transition model \[65\] of the form \(\sigma(n) = A(n-n_p)^{1.31}\), and extract the percolation threshold density \((n_p)\). \(n_p\) gives a measure of the minimum number of carriers required to fill the disorder landscape before a conducting pathway can be supported.
Disorder at the conduction band edge in the MOS system can be modelled as a 2D energy landscape with “hills” and “valleys” in energy (Figure 4.8). Conduction through this system can be modelled as a classical resistor network of saddle points, or quantum point contacts (QPC), where disorder is accounted for by the critical energy ($\epsilon_c$) or “height” of the saddle point \[99\]. Conduction $G(E_F, T)$ through each QPC is given by:

$$G(E_F, T) = \frac{2e^2}{h} \frac{1}{1 + \exp[(\epsilon_c - E_F)/k_BT]}$$  \hspace{0.5cm} (4.3)

At 0 temperature, the individual QPC conductance is a step function where $G = 2e^2/h$ for $E_F > \epsilon_c$ and $G = 0$ for $E_F < \epsilon_c$. For very low electron densities, $E_F$ lies below $\epsilon_c$ for the majority of QPCs in the network, so only a few isolated QPCs are conductive and as a whole the system is insulating. As the electron density is increased, $E_F$ rises and more and more QPCs become conductive. At a certain critical electron density, enough QPCs turn on such that a conductive pathway is formed from one end of the system to the other end. This critical density is known as the T=0 percolation threshold, $n_p(T = 0)$ or $n_0$. Above the percolation threshold density, the behavior of the conductivity $\sigma(n)$ is known to scale as:

$$\sigma(n) = A(n - n_p)^{1.31}$$  \hspace{0.5cm} (4.4)

The exponent 1.31 is the universal critical percolation exponent for two dimensions and is known from numerics \[65\]. The percolation threshold density is larger for a more disordered system.

From the transport data from devices A and B, we can extract the T=0 percolation threshold density by fitting the conductivity to expression \[4.4\] at each measured temperature and extrapolating the percolation threshold down to 0 temperature. Holding the critical percolation exponent $p$ at 1.31 \[65\], and fixing the pre-factor $A
Figure 4.9: Percolation threshold density ($n_p$) for samples A and B as a function of temperature. Fits to the form $n_p = n_0 + Ce^{-b/T}$ are also shown. Inset shows conductivity data ($\sigma$) vs. $n$ and percolation fits with $p = 1.31$ to extract $n_p$ for sample A at 335 mK and sample B at 345 mK.

To the best-fit value obtained for each device at the lowest temperature measured, we extract a value of $n_p$ at each measured temperature (Figure 4.9). The inset of Figure 4.9 plots the conductivity data of device A and B at 300 mK fit to the expression 4.4.

Figure 4.9 shows that (somewhat unintuitively) the measured percolation threshold increases with higher temperature. One may intuitively guess that with higher temperature the percolation threshold should decrease as the increase in thermal energy allows localized electrons to hop over energy barriers and contribute to conduction, yielding a lower percolation threshold. It is true that with higher temperature and at the lowest electron densities ($\approx 10^{11}$ cm$^{-2}$) hopping conduction dominates [65]. However, with higher temperature, electron screening of charged impurities becomes
less efficient as the polarizability of screening electrons smears out with temperature (the same mechanism that causes the peak mobility to be higher at 300 mK compared to 4 K) [76]. Therefore, due to the less efficient screening of Coulomb scattering centers at higher temperature, the extracted percolation density is higher with higher temperature, saturating at the T=0 percolation threshold at temperatures approximately below 1 K.

Using the functional form \( n_p = n_0 + Ce^{-b/T} \), we extrapolate the percolation threshold to zero temperature and extract \( n_0 \) [65]. The exponential term \( b \) is an energy gap related to the impurity distribution of the system. Our fit yields \( n_p = 0.83 + 1.46e^{-2.25/T} \) for sample A and \( n_p = 0.95 + 0.88e^{-3.00/T} \) for sample B, showing very similar T=0 percolation thresholds of \( 0.83 \times 10^{11} \) cm\(^{-2} \) and \( 0.95 \times 10^{11} \) cm\(^{-2} \), respectively.

This demonstrates that despite almost a factor of 2 difference in the peak mobilities of these two devices, the forming gas anneal is effective in restoring the device quality after e-beam irradiation, as measured by the percolation threshold density. In addition, these percolation threshold values approach (within a factor of 2-3) the percolation threshold values measured in Si/SiGe devices [32].

### 4.4.3 Electron Spin Resonance Measurements

Using ESR we now directly measured the density of electrons confined in shallow traps in samples A and B [60 [100]. We use X-band (\( \sim 9.6 \) GHz, \( \sim 3400 \) G) continuous wave ESR to measure the intensity of the 2DEG spin signal as a function of gate voltage (\( V_G \)) at fixed temperature between 360 mK and 4.2 K. Figure 4.10 shows an example of the number of unpaired spins, calculated as the double integral of the ESR spectrum, as a function of \( V_G \). The data shown in Figure 4.10 is from sample B, measured at 1.50 K.
Figure 4.10: Total number of unpaired spins plotted as a function of gate voltage for sample B at 1.50 K as measured in the dark (circles) and after above gap illumination (squares). $V_{th} = 0.14$ V at this temperature, shown by the dashed vertical line. Inset: ESR spectra at $V_G = 0.0$ V measured in the dark and after above gap illumination.

In this plot we can resolve three regimes. For $V_G$ above threshold conduction electrons fill the MOSFET channel and the ESR signal is roughly constant, reflecting the constant 2D density of states. As $V_G$ is scanned below threshold, electrons are localized into shallow traps and the channel is no longer conductive. The ESR signal decreases as $V_G$ is made more negative as electrons confined in shallow traps can thermally hop to the source and drain contacts. At some $V_G$, the ESR signal saturates (“dark” curve in Fig. 4.10) when the chemical potential is aligned with shallow traps deep enough that the confined electrons cannot thermally escape. These electrons are now frozen into shallow traps. We denote this characteristic voltage as $V^*$. Illuminating the sample with above band gap (1050 nm) light relaxes the system by
neutralizing confined electrons with holes and the corresponding (“post-LED”) ESR signal decreases and eventually goes to zero at voltage $V^0$.

With values for $V^*$ and $V^0$, we may then calculate the number of electrons confined in shallow traps ($n_{conf}$) at each measured temperature using the relation $e \cdot n_{conf} = C_{ox}(V^* - V^0)$, where $C_{ox}$ is the oxide capacitance measured from the Hall resistivity. The energy scale of the shallowest populated traps at each temperature [60, 101] can be calculated as approximately $10k_B \cdot T$. This estimate of the shallow trap depth comes from the Jenq method [102] in which the characteristic escape time $\tau$ of the trap is given by:

$$\frac{1}{\tau} = n v_{th} \sigma \exp\left(\frac{E_T - E_F}{k_B T}\right) \frac{1}{1 + \exp\left(\frac{E_T - E_F}{k_B T}\right)}$$  (4.5)

Here, $n$ is the density of conduction band electrons, $v_{th}$ is the thermal velocity of conduction band electrons, $\sigma$ is the capture cross-section (taken to be $10^{-19}$ m$^2$) and $E_T$ is the energy of the trap. The escape time is taken to be much longer than the timescale of the experiment ($10^5$ s) and then the trap depth $E_T$ can be numerically solved for as a function of temperature.

Figure 4.11 summarizes the measured shallow trap density as a function of temperature for samples A and B and plots data from a previous study for reference [60]. We note that the samples studied in the current work demonstrate similar densities of shallow traps to the previously studied Sandia device [60]. As temperature decreases, the density of confined charge increases as more electrons are frozen into shallower traps. Within experimental error, samples A and B demonstrate the same density of shallow traps across the measured temperature range, $\approx 9 \times 10^{10}$ cm$^{-2}$ at $\geq 0.3$ meV below $E_C$ (360 mK) and $\approx 3 \times 10^{10}$ cm$^{-2}$ at $\geq 2$ meV below $E_C$ (2.0 K). This measurement is consistent with the T=0 percolation threshold densities for both devices studied ($n_0^{A,B}$), also plotted in figure 4.11 and demonstrates the efficacy
Figure 4.11: ESR measurement of density of electrons confined in shallow traps as a function of temperature for samples A and B compared to previously studied devices (triangles). T=0 percolation threshold densities for A (hexagram) and B (pentagram), (from Fig. 4.9) are also plotted on the axis.

of the forming gas anneal for removing e-beam generated shallow defects across the measured energy range.

4.5 Conclusions

In summary, we have fabricated and measured high-mobility MOSFETs and have shown that a forming gas anneal is sufficient to restore an e-beam irradiated sample to the quality of an un-irradiated sample, as measured by the extraction of a T=0 percolation threshold and ESR measurements of the density of shallow traps. One of the devices measured in this study demonstrates the highest electron mobility for
a thin-oxide MOSFET. Furthermore, we demonstrate agreement between two independent methods of assessing the oxide interface, discovering a correlation between transport measurements of the T=0 percolation threshold density and ESR measurements of shallow electron traps. We believe that these measurements, as opposed to peak mobility, are more relevant metrics to characterize MOS interfaces for quantum devices operating in the low electron density regime and demonstrate a platform for understanding disorder relevant to quantum dot devices.
Chapter 5

Metal-Oxide-Semiconductor

Double Quantum Dots

In the previous chapters we have developed a fabrication process yielding very low densities of shallow traps in spite of processes known to generate high densities of defects. We now leverage this knowledge to fabricate a low disorder MOS double quantum dot device. Our fabrication process yields critical densities approaching the critical densities reported in Si/SiGe devices. We perform transport measurements of the device at 300 mK, utilizing a nearby charge sensor. We find that the device can be tuned to the single-electron regime where charging energies of \( \approx 8 \) meV are measured in both dots, consistent with the lithographic size of the dot. Additionally, we utilize the quantum dot and charge sensor as a local probe of nearby electron traps, finding three distinct traps that are constant from cooldown to cooldown. We find that this microscopic defect density is in agreement with the ensemble measurement of shallow traps measured in a bulk sample in the previous chapter. We measure the charge noise characteristics of the device which indicate a \( 1/f \) noise spectrum of 3.4 \( \mu \)eV/Hz\(^{1/2}\) at 1 Hz, consistent with other silicon quantum dot devices measured at
mK. Additionally, magnetospectroscopy measurements yield a valley splitting of 110±26 µeV. These experiments have been published in reference [103].

With these measurements, we have correlated for the first time mesoscopic defect measurements with ensemble measurements of defect densities, demonstrating a method of determining high quality materials for quantum dot devices. In addition, while other groups have resorted to fabricating very small (~25 nm radius) quantum dots in an attempt to avoid defects [55, 24, 19], we have demonstrated quantum dots nearly twice as large with low levels of disorder and comparable charge noise [104] and valley splittings [54]. With this improvement in defect density, these devices pave a way for scaling to larger MOS quantum dot systems.

5.1 Introduction

Electron spins in silicon devices are promising qubits for a quantum processor, defining a natural two-level system and demonstrating long spin coherence times [14, 15, 105]. Recently, 2-qubit quantum controlled-not (CNOT) operations have been demonstrated in both Si/SiGe [29] and Si Metal-Oxide-Semiconductor (MOS) quantum dot systems [28], establishing a crucial building block for a universal quantum computer [7]. While the MOS system allows for electron-donor interactions (enabling an avenue for quantum memories in donor states) [106, 27] and demonstrates larger valley splittings (critical for high fidelity spin-selective operations) [107, 54, 35, 32, 56], it suffers from high disorder compared to its Si/SiGe counterpart [13]. Indeed, because of disorder, the scaling-up of MOS quantum dots to multi-qubit systems has lagged behind Si/SiGe systems, where an array of nine uniform quantum dots have been achieved [108] as well as the coupling of a single spin to a superconducting resonator [109, 110]. Furthermore, in MOS, disorder may be unintentionally introduced to the Si/SiO₂ interface during high-energy processes like electron-beam lithography [21, 20],
an essential fabrication process for quantum dot devices produced by research labs. In this work we address the issue of disorder in MOS quantum dots by engineering and characterizing a low-disorder MOS quantum dot device where the disorder parameters critical for quantum dot devices, i.e. density of shallow traps and critical density \[21\, 60\], approach the low-disorder levels demonstrated in Si/SiGe systems \[32\, 64\].

We have fabricated and characterized a low disorder MOS quantum double-dot device, leveraging the fabrication process detailed in the previous chapter, yielding very low critical and shallow trap densities \( (8.3 - 9.5 \times 10^{10} \text{ cm}^{-2}) \), and simultaneously very high mobilities \( (1.4 - 2.3 \times 10^4 \text{ cm}^2/\text{Vs}) \), despite exposure to high-energy processes like electron-beam lithography \[21\]. We note that the critical densities of this starting gate stack are within a factor of 2-3 of critical densities reported in Si/SiGe devices \( (4.6 \times 10^{10} \text{ cm}^{-2}) \) \[32\] and are on par with the lowest critical densities reported in MOS \[63\].

We adopt a reconfigurable device architecture pioneered in Si/SiGe devices \[32\], with three overlapping layers of gates defining two parallel conduction channels. Notably, our device’s first layer of gates are fabricated from degenerately doped poly-silicon instead of the typical aluminum. Low frequency bias spectroscopy measurements through each individual dot in the upper channel show regular Coulomb blockade diamonds, demonstrating low levels of disorder and the charging of a single lithographically defined dot. We define a charge sensor dot in the center of the lower conduction channel and detect the depletion of each upper channel quantum dot to the single-electron regime and additionally show the controllable formation of a quantum double-dot. Our charge sensor also detects the charging of individual poly-silicon grains (“phantom dots”) within the poly-silicon depletion gates which do not affect the charge state of our quantum dots but add a prominent background signal.
Figure 5.1: Basic fabrication flow diagram of double dot device.

Charge-noise spectroscopy measurements yield a $1/f$ spectrum in power spectral density with a value of $3.4 \, \mu\text{eV}/\text{Hz}^{1/2}$ at 1 Hz, consistent with other reported values of charge noise in MOS and Si/SiGe devices [104]. Monitoring the $N=0 \rightarrow 1$ and $N=1 \rightarrow 2$ transition as a function of a perpendicular magnetic field, we measure a valley splitting of $110 \pm 26 \, \mu\text{eV}$.

5.2 Device Fabrication

The device fabrication process was engineered to minimize shallow traps by minimizing high-energy processes and eliminating vectors of mobile ionic contamination. Starting from the MOS gate stack detailed in the previous chapters, we adopt an overlapping gate architecture [32] capable of defining up to four individual quantum dots with independent control of each dot’s electron occupation, tunnel barriers and source/drain reservoirs. Figure 5.1 illustrate the fabrication process flow. A $300 \times 300 \, \mu\text{m}^2$ window is implanted with arsenic which defines the device area. The polysilicon
Figure 5.2: (a) False color SEM micrograph of a device identical to the one measured. Schematic of (b) horizontal and (c) vertical cross-sections of the device as indicated by the dotted lines in (a).

gate is degenerately doped in this step as are four ohmic contacts in the silicon. The device is then annealed at high temperature (900 C) to activate the implant.

The first layer of gates is achieved by thinning the poly-Si gate down to 40 nm with a low power reactive ion etch step (50 W, SF₆ and C₄F₈) and etching two parallel 80 nm wide channels, defined by electron-beam (e-beam) lithography (Elionix F-125), down to the SiO₂. These channels define three poly-Si depletion gates (labelled Poly₁,₂,₃ in Fig. 5.2(a)) which serve as a screening layer for the subsequent Al accumulation gates and also prevent the diffusion of mobile ionic contaminants into the underlying gate oxide during the fabrication process. Immediately after stripping the e-beam resist, 200 cycles (22 nm) of atomic layer deposition aluminum oxide (ALD) were grown on the sample to insulate the poly-Si gates from the subsequent Al gates and to protect the exposed oxide from contamination. Two subsequent layers of 50 nm wide aluminum gates were deposited via lift-off, defining source/drain reservoirs (labelled Source₁/₂, Drain₁/₂), plunger gates (labelled PL₁/₂, PR₁/₂), and
Figure 5.3: Optical image of a quantum dot device bonded to a PCB.

tunnel barrier gates (labelled TL1/2, TM1/2, TR1/2). These gates were oxidized between depositions in an O2 plasma for electrical insulation [95]. Finally the device was furnace annealed at 400C in forming gas (95% N2, 5% H2) for 30 minutes to lower the interface state density and repair damage incurred from the e-beam [21, 67].

Each of the device gates extends out to the edge of the chip to an aluminum bond pad where the underlying polysilicon is un-doped. The device is mounted to PCB with Elmer’s rubber cement and wire-bonded with a wedge bonder (Figure 5.3).

One notable change in the materials in this gate stack and the low-disorder starting gate stack is the addition of ALD aluminum oxide on top of the thermal gate SiO2. The introduction of new materials to the gate stack can alter the quality of the Si/SiO2 interface as the SiO2/ALD interface is known to contain substantial amounts of fixed negative charge [111]. However, Van der Pauw devices fabricated with this new Si/SiO2/ALD/Al gate stack achieved low electron densities ($\sim 1.0 \times 10^{11} \text{ cm}^{-2}$).
at 4.2 K) and high mobilities (\(1.4 \times 10^4\) cm\(^2\)/Vs) so we are confident that the effect of the ALD on the Si/SiO\(_2\) interface disorder is minimal (Figure 5.4).

### 5.3 Experimental Setup

This device was measured in a \(^3\)He cryostat where low frequency transport measurements were taken using an Ithaco 1211 transimpedance amplifier in conjunction with an SR830 lock-in amplifier at a frequency of 229 Hz and an AC excitation amplitude of 50 \(\mu\)V. The AC excitation amplitude is achieved through a 10,000:1 voltage divider. A schematic of the instrumentation set-up is shown in Figure 5.5 and a photograph of the set-up is shown in Figure 5.6. A simple bias tee circuit was implemented with the AC modulation in order to add a DC bias voltage underneath the AC modulation. Nine SRS SIM928 low noise voltage sources are used to supply DC voltages to the gates on the device. These voltage sources can supply up to 20 V with 1 mV resolu-
tion, but are generally limited to 4 V in order to prevent leakage between gates. For increased resolution, we sometimes utilize a 5:1 voltage divider. The SRS 830 also features four auxiliary voltage outputs that are sometimes used to supply voltages to the device. The auxiliary voltage outputs are low-pass filtered at room temperature at 6 Hz because they can apply a noisy transient to the device when the SRS 830 is communicating with the PC. For noise spectroscopy measurements, the output of the Ithaco 1211 is fed instead to an HP3561a signal analyzer. For magnetospectroscopy measurements, we energize a Cryomagnetics superconducting magnet up to 4 T. Each instrument is interfaced to the measurement PC through an optical USB hub to break ground loops from the measurement set-up to the PC.

Inside of the $^3$He cryostat sample chamber, we installed a home-made silver epoxy RF filter [112] on the wires connected to our device. The filter consists of $\approx 1$ m of constantan loom wrapped around a copper post and affixed with colloidal silver epoxy. The silver epoxy is a resistive metal which can dissipate RF noise travelling
Figure 5.6: Photograph of instrumentation set up for quantum dot transport measurements.

down the electrical lines through eddy currents. The filter is screwed into the $^3$He cold plate of the cryostat and additionally thermally anchors the device to the $^3$He pot. With this filter, our electron temperature improved from $\approx 1$ K to the base temperature of the cryostat, 300 mK.

Ground loops in the system can add significant 60 Hz noise to the measurement lines and can easily increase the electron temperature. We utilize a true earth ground as our measurement ground and ground all instruments and cable shielding in a star arrangement radiating out from the single ground source to the best of our ability. The cryostat is also grounded to the measurement ground. Within the system, several major sources of 60 Hz noise were found and corrected. A major 60 Hz ground loop was found in the bias tee circuit. Isolating the bias tee shielding from the AC output shielding and the DC output shielding significantly reduced this 60 Hz noise. Another major 60 Hz source came from a temperature controller that was connected to four wires in the loom connecting to the device. Grounding these four wires and disconnecting the temperature controller fixed this issue. Any stray, unused wires within the cryostat should be grounded as well. Any RC filters and voltage dividers
in the system need to be grounded carefully as these components can also introduce ground loops. In addition, any resistors used in any circuit components should be metal film instead of carbon to reduce resistor noise. Ground isolated Pomona boxes are especially useful in finding and breaking ground loops as they can be easily inserted into a suspected ground loop. 60 Hz noise can easily be measured by feeding the lock-in signal monitor to an oscilloscope and triggering it to 60 Hz, or by feeding the signal monitor to a spectrum analyzer.

One final instrumentation note pertains to electro-static discharge (ESD). Due to the small gate size of our quantum dot device, these devices are very susceptible to ESD, so the electrical connections to the device should either all ideally connected to a controlled voltage source or ground or, less ideally, shorted all together (when the device is unplugged from the system). The breakout box interfacing the measurement instruments to the device is thus designed to seamlessly connect a voltage source to the device, or connect the device to ground so that BNC connections on the box can be safely swapped (Figure 5.7). The breakout box is equipped with two sets of redundant switches for each electrical line. The first is a make-before-break (MBB)
rotary switch that either connects the device line to ground or to a BNC connector on
the face of the box (where a voltage source can be plugged in). Each MBB will switch
6 device lines to 6 different BNC connections simultaneously. As the name suggest,
the new connection is made before the old connection is broken so that the device is
never floating. The second switch is a single-pole single-throw (SPST) switch that
will short the BNC connector to ground. Thus, to safely swap the BNC connections,
the voltage source is first set to 0, then one of the switches is thrown to short the
device connection to ground, then the BNC cable is detached. If the switches are used
correctly, the device leads will never be floating. When the device is being unplugged
from the system, the PCB leads connected to the device are shorted together with a
piece of aluminum tape.

5.4 Tuning a Quantum Dot

The device is first screened by mounting it to a dipper probe and dipping it into a
liquid helium dewar. Each gate is tested to ensure it is coupled to the 2DEG and
can pinch off current through the quantum dot channel. The device is then mounted
in the sample chamber of a $^3$He cryostat and cooled down. Typically, at 4 K (before
cooling to base temperature) a we flash a 1050 nm LED to neutralize any electrons
confined in shallow traps.

Transport is measured through the channel by raising each gate to about 3 V.
The polysilicon depletion gates are typically held at 0 V. Then each gate voltage is
individually lowered until the channel current pinches off. Usually resonances are
observed in the pinch-off curves. These are probably associated with disorder at the
interface which can usually be tuned away. These pinch off voltages are recorded for
each gate (Figure 5.8).
Figure 5.8: Pinch-off curves for all five gates associated with the upper channel of the double quantum dot device. Each gate is held at 3 V while an individual gate is scanned down. Each gate pinches off between 1 and 1.5 V.

Figure 5.9: 2D conductance scan of tunnel barriers (UTM, UTR) adjacent to a QD. The parallel conductance lines converge to a “corner” at (0.8,0.7) V.
To tune an individual quantum dot now, we set the plunger voltage (UPR) somewhere high above its pinch-off voltage and perform a 2D scan of the two adjacent tunnel barriers (UTM and UTR in Figure 5.9). As the tunnel barrier voltages are scanned down, tunnel barriers are formed around the plunger, defining a dot. In this measurement, symmetric diagonal oscillations in the conductance signal are observed converging to a “corner” where conduction is bounded by the edges of the corner. If either tunnel barrier voltage becomes too low, conduction through the dot is completely pinched off. The diagonal conductance oscillations are Coulomb blockade oscillations that are being brought into resonance by the cross-capacitance of the tunnel gates to the quantum dot. In this regime, a single, well defined quantum dot is formed such that both tunnel barriers are equally coupled to the dot. The tunnel barrier voltages are then set somewhere near the corner and then the plunger voltage is scanned yielding the typical Coulomb blockade oscillation picture (Figure 5.10).
5.5 Bias Spectroscopy

For this work, we bias the device to form two quantum dots in the upper conduction channel and define a charge sensor dot in the center of the lower conduction channel. The electron temperature for these experiments is $\sim$300 mK, estimated by the line width of the Coulomb blockade peaks.
We first characterize each quantum dot individually with low-frequency bias spectroscopy measurements to extract the charging energies and lever arms of each dot. We observe regular Coulomb blockade diamonds over a wide electron occupation range which monotonically decrease in size with increasing electron occupation, reflecting the increase in quantum dot size with increasing voltage applied to the plunger gate (Fig. 5.11 (a)). At lower plunger voltages, the Coulomb diamonds appear to cease, indicating the dot has reached the few-electron regime. In this regime, the charging energies ($E_C$) for both dots are measured to be $\approx 5$ meV, and we extract lever arms of 0.059 meV/mV for the right dot and 0.067 meV/mV for the left dot, respectively (Fig. 5.11(b), (c)).

### 5.6 Charge Sensing Measurements

As the plunger voltage is made more and more negative to depopulate the quantum dot, the tunnel barriers simultaneously become more opaque due to the cross capacitance of the plunger gate to the tunnel barriers. Thus, as the quantum dot is depopulated, the tunnel rate through the dot decreases, resulting in a conductance signal which falls below the noise floor of the measurement system. In order to interrogate the single-electron regime, we define a charge sensor dot in the center of the lower conduction channel beneath TM$_2$. When the charge sensor dot is biased to the edge of a Coulomb blockade peak, small changes to the local electrostatic environment (e.g. the addition of single electrons in the upper quantum dots) result in a measurable change in current through the charge sensor dot. With the charge sensor, we may explicitly show that the upper quantum dots have reached the single-electron regime.

Figure 5.13 shows the charge stability diagrams measured by the charge sensor dot of the the upper two quantum dots as a function of the corresponding quantum dot...
5.6.1 Quantum Dot Transitions

The set of QD transitions are strongly coupled to the plunger gate voltages and their charging energies are consistent with the charging energy measured by bias spectroscopy. In addition, they show regularity over a large number of electron transitions, indicative of low interface disorder. As the tunnel barrier is made more negative (more opaque), the QD confinement increases and thus the spacing between QD
Figure 5.13: Charge stability diagrams for the upper right dot (a), (b) and upper left dot (c), demonstrating the tuning of each dot down to 0 electrons. The regular parallel black lines indicate the quantum dot transitions. Three electron traps in the vicinity of the quantum dots are detected and are highlighted by the red dotted lines in (a)-(c). A persistent background oscillation is also present in the charge-sensing signal from the charging of individual poly-silicon grains in the poly-silicon depletion gate. (d) demonstrates the controllable formation of a quantum double dot.
transitions increases slightly. At sufficiently negative tunnel gate voltages, the tunnel barriers increase in opacity to the point where the tunnel rate can no longer keep up with the measurement scan rate and “latching” behavior is seen in the lower region of the left-most transitions. Monitoring the transition point of the latching behavior for each electron transition ensures that the tunnel rate of each electron transition remains fast enough to be detected during the scan. In this way, electron transitions are not obscured by a tunnel rate being much slower than the measurement scan rate. Thus we demonstrate the depletion of each quantum dot to zero electrons.

Utilizing the charge sensor we may now extract the charging energies for the first electron transition, converting the plunger gate voltage difference between the first and second electron transitions to energy using the lever arms measured from the Coulomb blockade diamonds (Fig. 5.11). We extract charging energies of 7.9 meV for the right dot and 7.6 meV for the left dot. Treating each dot as a metallic disc at the Si/SiO$_2$ interface, we can estimate the radius (R) of each dot in the single-electron regime from the dot’s measured charging energy ($E_C$) and calculated self capacitance:

$$E_C = \frac{e^2}{C_{\text{disc}}}$$

where $C_{\text{disc}} = 8\bar{\epsilon}\epsilon_0 R$ [32]. Here, $\bar{\epsilon}$ is given by $\frac{1}{2}(\epsilon_{\text{Si}} + \epsilon_{\text{ox}})$, the arithmetic mean of the relative permittivity of silicon ($\epsilon_{\text{Si}}=11.7$) and silicon dioxide ($\epsilon_{\text{ox}}=3.9$) [36]. This analysis yields radii of 37 nm for the right dot and 38 nm for the left dot, in good agreement with the effective lithographic radius of the dots (50x80nm$^2 \rightarrow R=35.7$nm).

In addition, bias-spectroscopy measurements of the left dot in the single-electron regime (Figure 5.14) reveal a conductance resonance 1.2 meV above the ground state which we attribute to the first excited orbital state. Treating the dot now as a 2D square box [32], we estimate the energy of the first excited state as $E_{\text{orb}} = \frac{3\hbar^2\pi^2}{2m^*L^2}$, where $m^* = 0.19m_0$ and $L$ is taken to be $(\pi R^2)^{1/2}$. This analysis yields $E_{\text{orb}} = 1.3$ meV, in good agreement with the measured resonance. From these data, we conclude
Figure 5.14: Bias spectroscopy plot of the first electron transition in the left dot. The ground state and first excited state are indicated by red arrows.

that these dots are lithographically defined and not dominated by random disorder at the Si/SiO$_2$ interface.

5.6.2 “Phantom Dot” Transitions

In the background of the charge sensing signal is a persistent signal caused by the charging of individual grains in the poly-silicon depletion gate. We determine these charging events to originate from the poly-silicon layer instead of the Si/SiO$_2$ interface disorder by three pieces of evidence. First, the transitions persist for all gate voltages probed and never deplete, indicating the charging object is a metallic island instead of a semiconductor dot. Second, the charging events are only evident in the charge
sensing signal and not in the bias spectroscopy signal, indicating the object being charged does not originate from the Si/SiO$_2$ interface. The third and strongest piece of evidence is that the slope of these charging events turns positive when a charge sensing measurement is performed scanning the voltage of the poly-silicon gate against any other gate. A positive slope of these transitions is only possible when the object being charged is present on one of the gates being energized during the measurement. In this situation, the poly-silicon gate acts as both the electron reservoir and plunger gate to a poly-silicon grain acting as a quantum dot. Thus the Fermi level of the poly-silicon grain is pinned to the Fermi level of the surrounding poly-silicon gate. We emphasize that the poly-silicon transitions do not interact with or otherwise affect the charge transitions of the quantum dots, but simply add a noisy background signal. These “phantom dot” signatures have been observed in other similar devices and can be eliminated by keeping the thickness of the poly-silicon layer thicker than the order of the poly-silicon grain size.

5.6.3 Defect Detection

The charge sensor in conjunction with the quantum dot can also be used as a local probe of defect states in the vicinity of the quantum dot. Defect states are evidenced by a single transition line slanted away from the set of parallel quantum dot transitions, highlighted in Fig. 5.13 (a)-(c), creating avoided crossing with the QD transitions. Scanning the QD gate voltages over a wide parameter range (limited to 4V to avoid leakage between gates) to search for defects, we identify three distinct defect states, two in the vicinity of the right dot (highlighted in Fig. 5.13 (a) and (b)) and one in the vicinity of the left dot (highlighted in Fig. 5.13 (c)). These defect states are observed to hold only a single electron and the location of the defect state transition relative to the QD transitions are consistent from cool down to cool down. These observations suggest that the origin of these defects is a fixed positive charge
in the oxide near the interface, as opposed to a mobile ionic charge which can freely migrate through the oxide at room temperature [67]. Using the lithographic size of the dots, a rough estimate for the defect density can be obtained yielding $\approx 3 \times 10^{10}$ cm$^{-2}$. The estimate of the defect density from this method is consistent with the order of magnitude of the defect density measured by ensemble electron spin resonance measurements of the shallow trap density and conductivity measurements of the critical density of this material [21].

5.6.4 Double Dot

Finally, we demonstrate the controllable formation of a quantum double dot (Fig. 5.13 (d)). We observe the classic “honey-comb” structure and triple-points. In this regime, there are several electrons in each quantum dot. We note that this double quantum dot enables a promising platform for future two-qubit operations.

5.7 1/$f$ Charge Noise

Another important device parameter is charge noise characteristic. By biasing one of the dots to the edge of a Coulomb blockade peak (Figure 5.15), the dot is maximally sensitive to charge noise caused by fluctuations in the local electric field environment. In this measurement, a 1 mV DC bias is applied to the source of the device and the drain current is measured by the current preamplifier. The noise spectrum is then obtained from the output of the current preamplifier by an HP3561a signal analyzer. By comparing noise spectrum data at the edge of the Coulomb blockade peak with the noise spectrum taken at the top of the Coulomb blockade peak (where the device is minimally sensitive to local noise), the noise generated by the local environment can be extracted from noise generated by all other noise sources in the device and measurement circuitry [104, 113].
Figure 5.15: Quantum dot drain current as a function of plunger voltage with a 1 mV bias across the source/drain leads. The red square on the edge of the Coulomb blockade peak is where the sensor is maximally sensitive to local variations in the charge noise environment. The blue star indicates where the sensor is minimally sensitive to variations in the charge noise environment.

Figure 5.16: (a) Derivative of the QD drain current across a Coulomb blockade peak (left axis) overlaid with the power spectral density for 0.5, 1.0, 1.5 Hz (right axis). (b) Measured power spectral density at the maximally sensitive and minimally sensitive points of the QD Coulomb blockade peak (star and box indicated in (a)) as a function of frequency. $1/f$ dashed line is shown for comparison.
Figure 5.16(a) shows the derivative of the quantum dot drain current with respect to the plunger voltage \( \frac{dI_{SD}}{dV_{PR1}} \) across a Coulomb blockade peak. Overlaid on the same plot is the magnitude of the noise spectrum at 0.5, 1.0, and 1.5 Hz. As expected, the magnitude of the noise correlates with the absolute value of derivative of the device current, indicating the noise spectra are dominated by the fluctuations in the dot’s chemical potential, \( \epsilon \), from local environmental noise sources [104]. The noise spectra at the maximum and minimum points of \( \frac{dI_{SD}}{dV_{PR1}} \), (indicated in Fig. 5.16(a) by \( V_{S_{\text{max}}} \) and \( V_{S_{\text{min}}} \)) are plotted in Fig. 5.16(b) from 0.2 Hz to 59 Hz. The noise spectra follow a low frequency \( 1/f \) dependence, frequently observed in electronic devices and indicative of an ensemble of two-level fluctuators in the vicinity of the device [114]. The measured magnitude of the noise spectra is well above the measured white-noise floor of our current preamplifier.

We convert the measured current noise to the equivalent potential noise felt by the quantum dot, \( \Delta \epsilon \), using the relation \( \Delta I_\epsilon \alpha = \left| \frac{dI_{sd}}{dV_{PR1}} \right| \Delta \epsilon \) [104, 113]. Here, \( \Delta I_\epsilon = \sqrt{S_{\text{max}} - S_{\text{min}}} \). For \( \Delta \epsilon \) evaluated at 1 Hz, we calculate a noise value of 3.4 \( \mu \text{eV/Hz}^{1/2} \), consistent with other reported noise values reported at 1 Hz at 300 mK in MOS and Si/SiGe quantum devices [104]. We note the temperature dependence of \( 1/f \) noise in semiconductor quantum devices has been observed to decrease with temperature [115, 104] so our measured noise figure is likely to decrease at dilution refrigerator temperatures.

### 5.8 Valley Splitting

Finally, we perform magneto-spectroscopy measurements of the \( N=0 \rightarrow 1 \) and \( N=1 \rightarrow 2 \) electron transitions to measure the valley splitting. In bulk silicon, electrons reside in one of six degenerate valley states corresponding to the six symmetric minimums of the conduction band edge in \( k \)-space. When electrons are confined to an interface
in a quantum dot, this six-fold degeneracy is lifted such that electrons preferentially populate the two valleys perpendicular to the interface [52]. The degeneracy of these remaining two valleys is referred to as the valley splitting and can be lifted through electric fields, confinement, and other details of the interface [52, 54]. For high fidelity spin selective quantum operations, the valley splitting must be large in relation to $k_B T$ and the qubit energy [52].

Figure 5.17(c) illustrates the energy diagram of the first two electrons loading into the quantum dot as a function of the magnetic field. The first electron loads into the lower valley state in the spin down configuration, and its addition energy decreases as a function of the magnetic field with a slope of $-\frac{1}{2} g\mu_B B = -58 \, \mu eV/T$, where the $\mu_B$ is the Bohr magneton and the $g$-factor is taken to be 2. The second electron also loads into the dot in the lower valley state, but in the spin up configuration at low magnetic field, forming a spin singlet ground state, and its addition energy increases with a slope of $+\frac{1}{2} g\mu_B$. At a certain magnetic field value, the spin up arm of the
lower valley state crosses the spin down arm of the upper valley state. Above this magnetic field, the second electron loads into a spin down configuration of the upper valley state, forming a spin triplet with the first electron. By subtracting the addition energy of the first electron from the second electron, we compensate for any effects of the magnetic field on the quantum dot orbital states [56] (which can be significant in a perpendicular magnetic field) as well as any charge offset drift [116, 117] (i.e. very low frequency drifts in the electrostatic environment on the order of 0.01 e). The difference in the transition energies of the first and second electron should first increase with a slope of $g\mu_B$ in the spin singlet configuration and then should be flat in the spin triplet configuration.

In a perpendicular magnetic field, we tune the upper left dot to the first and second electron transitions, then monitor the position of each transition’s Coulomb blockade peak as a function of magnetic field by fitting the conductance peak to a $\cosh^{-2}$ function [56]. We convert the peak position from the plunger voltage to energy using the measured lever arm. In this experiment, we measure the conductance of the quantum dot instead of utilizing our charge sensor in order to bypass the noisy polysilicon charging events which are apparent in the charge sensing signal. Figure 5.17(a) plots the evolution of the conductance peaks of the first two electron transitions as a function of magnetic field. The evolution of these peaks contains contributions from the orbital and spin energies (which are both magnetic field dependent) as well as charge offset drift (which is independent of magnetic field but constant for both electron transitions). By subtracting the position of the conductance peaks from each other, we may subtract out the orbital and charge offset drift components, leaving the difference in spin energies of the two electrons which is used to determine the valley splitting.

Figure 5.17(c) shows the difference in addition energies of the first and second electrons. The energy difference increases initially with a slope $g\mu_B$, then saturates at a
value of 110±26 µeV, corresponding to the electron transition to a spin triplet in the upper valley state. Thus, we demonstrate a valley splitting large enough in this device to support spin-selective operations at typical dilution refrigerator temperatures (∼100 mK = 8.7 µeV) [52].

5.9 Conclusions

In conclusion, we have engineered a low-disorder MOS quantum dot device and demonstrated a promising platform for electron spin qubits. We demonstrate the controllable formation of lithographically defined individual and double quantum dots with charging energies consistent with the lithographic size of the dots. The local defect density around the quantum dots is low enough to support single electron occupation and correlates with ensemble measurements of the defect density as measured by ESR and percolation thresholds. Charge noise spectroscopy measurements show a 1/f power spectral density yielding a value of 3.4 µeV/Hz^{1/2}, comparable to other Si MOS and Si/SiGe devices measured at 300 mK, and demonstrating a quiet noise environment. Finally, we measure a valley splitting of 110±26 µeV, large enough to support high-fidelity spin operations. This work represents a platform for optimizing quantum dot disorder in MOS and a promising architecture for spin qubits.
Chapter 6

Conclusion

In this thesis we have laid the groundwork for understanding and minimizing disorder in silicon MOS quantum devices and have demonstrated a promising device architecture for low-disorder MOS quantum dots. We reviewed the basics of quantum dot physics, the basics of MOS physics as well as the commonly accepted types of Si/SiO$_2$ interface defects. The defects that have been previously studied in the MOS literature turn out not to be the most relevant defect for quantum dot devices. The most detrimental, and least studied, defect for quantum dot devices are shallow electron traps, electron traps that reside within a few meV from the conduction band edge.

We explored device processing parameters that optimize the Si/SiO$_2$ interface quality, using the low temperature electron mobility as a proxy for interface quality. We develop a process that yields the highest mobility thin-oxide MOSFET, 23,000 cm$^2$/Vs at 300 mK. We find that low power reactive ion etching does minimal damage to the Si/SiO$_2$ interface. More importantly, trace sodium contamination can kill a device’s peak mobility, so precautions must be taken when the SiO$_2$ is exposed during high temperature and wet processes. Sodium contamination in our devices resulted a factor of 2 reduction in peak mobility.
A surprising result from exploring our device processing parameters is that the low temperature mobility experiences a significant boost when the device is annealed with aluminum covering the polysilicon gate. In devices with no aluminum present over the gate, the peak mobility is \( \approx 1,000 \text{ cm}^2/\text{Vs} \). During the forming gas anneal, aluminum cracks molecular hydrogen into atomic H which is extremely efficient at bonding to dangling bonds at the Si/SiO\(_2\) interface.

The low temperature electron mobility is an incomplete measure of the interface quality for quantum dot devices. In fact in a pair of previously studied devices with similar peak mobilities, a large difference in shallow trap densities was measured. We utilize electron spin resonance to directly measure the density of shallow electron traps in our high mobility devices and measure a low shallow trap density. Furthermore, we find that conventional forming gas annealing is sufficient to anneal out shallow electron traps generated by electron-beam lithography.

Finally in this experiment we demonstrate that transport measurements of the Si/SiO\(_2\) interface are correlated to the measured shallow trap density via the percolation threshold density. The percolation threshold density is backed out by extrapolating the minimum electron density required to support a conductive pathway in the sample. The percolation threshold extrapolated to 0 temperature agrees with the shallow trap density at the lowest temperature measured (\( \approx 8 - 9 \times 10^{10} \text{ cm}^{-2} \)). This lends itself to a consistent physical picture. Below threshold, electrons are localized in shallow traps. As the electron density is raised, electrons populate shallower and shallower traps until enough traps are filled to support a conducting pathway. This density correlates with the percolation threshold density measured by transport. The percolation threshold densities measured in these devices turn out to be on par with the best percolation (or critical) densities measured in MOS and are approaching the critical densities measured in Si/SiGe quantum dot devices.
Leveraging this base process yielding very high quality Si/SiO\textsubscript{2} interfaces, we fabricated and characterized a MOS double quantum dot device, adopting the dual rail architecture pioneered in Si/SiGe devices. A necessary modification in the device material stack is the inclusion of ALD aluminum oxide on top of the SiO\textsubscript{2}. ALD is known to contain substantial amounts of negative charge so it was unclear what effect it would have on the quality of the Si/SiO\textsubscript{2} interface. Test devices were fabricated with this new gate stack which yielded high mobilities (14,000 cm\textsuperscript{2}/Vs) and achieved low electron densities ($\sim 1 \times 10^{11}$ cm\textsuperscript{-2} at 4.2 K) so we are confident that the ALD does not introduce much disorder to the interface.

We cooled down and characterized our quantum dot devices in a $^{3}$He cryostat. The device demonstrates very regular blockade oscillations over a large parameter range, indicative of charging single, well-defined quantum dots. We measure charging energies of $\approx 8$ meV which is consistent with the lithographic size of our dots, treating the dots as metallic discs. We therefore conclude that our dots are lithographically defined and not dominated by random interface disorder.

Using an adjacent charge sensor dot, we can explicitly demonstrate that our dots have reached the single electron regime, realizing an important requirement for quantum operations. With the charge sensor and quantum dot, we can locally probe the area immediately surround the quantum dots for electron traps. We identify three distinct traps that are repeatable from cooldown to cooldown. A rough estimate of the defect density yields a shallow trap density on order of $3 \times 10^{10}$ cm\textsuperscript{-2} which agrees with the ensemble measured shallow trap density and percolation threshold.

We additionally measure the device’s charge noise characteristics. We measure a $1/f$ low frequency power spectral density which is commonly observed in electronic devices. By comparing the charge noise at the edge of a Coulomb blockade peak and the top of the Coulomb blockade peak we can extract the charge noise originating from the local electrostatic environment from the instrumental noise. We measure a
power spectral density of 3.4 µeV/Hz$^{1/2}$ which is consistent with other silicon devices measured at 300 mK.

Finally, we measure the valley splitting of our dots, tuning to the first and second electron transitions. We observe a singlet-triplet crossing near 1 T, corresponding to a valley splitting of 110±26 µeV (or 1.3 K), large enough to support high-fidelity spin operations.

\section{6.1 Future Work}

Moving forward, quantum dot systems will scale to larger and more complicated schemes. The current defect densities that have been achieved are workable with small scale quantum dot systems but must be improved in order to implement algorithms on large scale quantum dot systems.

One proposed source of shallow traps originates from the difference in work functions of different crystal grains in the gate material. In this work we have utilized polysilicon and aluminum gates, both of which are polycrystalline and have grain sizes on the order of 10’s of nanometers. The different facets of crystal grains are known to have different work functions. In a polycrystalline material, adjacent grains are randomly oriented from each other and as such will have different work functions at the SiO$_2$/gate interface. These differences in work functions result in a local electric field at the grain boundary which can translate to shallow electron traps at the Si/SiO$_2$ interface.

Utilizing an amorphous metal could lower the density of shallow traps as the work function at the surface of an amorphous material averages out any orientation dependencies. We are currently investigating an alloy of TaWSi, an amorphous metal, as a novel gate material. We are fabricating large-area MOSFETs with a Ta$_{40}$W$_{40}$Si$_{20}$ gate to measure the shallow trap density with ESR. A large reduction in shallow traps
would enable the scaling of large-scale, defect free quantum dot systems and motivate the use of amorphous metal gates. This effect would benefit both MOS and Si/SiGe systems.

Another important challenge in quantum dot architectures is to be able to shuttle individual electron spins from one quantum dot to another, nonadjacent quantum dot while preserving spin coherence. This will enable interactions beyond nearest neighbors and would reduce the number of physical qubits required in a universal quantum computer. With our low defect densities, it may be possible to shuttle a single electron in a quantum dot to another quantum dot on order of 1 \( \mu \text{m} \) away with the goal of maintaining its spin coherence.

We propose defining a micron long channel from one quantum dot to a second quantum dot and using a resistive metal like NbSi to drive a lateral electric field from one dot to the other. Each quantum dot would have an associate charge sensor dot for single shot readout. A first experiment would be to detect the shuttling of an individual electron from one dot to the other. An electron would be held in the first dot with a high tunnel barrier separating the dot from the channel. Current would then be pushed through the channel gate to induce a lateral field between the two dots. Then the tunnel barrier would be opened like a turnstile to allow the electron into the channel where it is shuttled to the next dot and detected. Subsequent measurements can initialize the electron such that relaxation and coherence measurements can be performed after the electron is shuttled. These measurements would have to be done at dilution refrigerator temperatures.

In summary, we have examined the defects most relevant for silicon quantum dot devices, demonstrating that standard annealing procedures are effective at minimizing shallow traps. We demonstrate that our ESR measurement of shallow traps agrees with the \( T=0 \) percolation threshold measured by transport, which gives us insight into the nature of these traps. We finally leverage these insights to fabricate a low-
disorder quantum double dot device which represents a promising architecture for future quantum dot physics in MOS.
Appendix A

Device Fabrication Details

A.1 Notes on Contamination

High mobility MOSFETs are incredibly sensitive to contamination, primarily from alkali ions like lithium, potassium, and sodium. These ions will freely diffuse through the silicon dioxide and will reduce the electron mobility and create charge traps at the Si/SiO$_2$ interface. As such, extreme care must be taken during device fabrication to avoid contaminating your devices.

- A huge source of sodium contamination comes from humans themselves, notably from finger grease. As such, do not handle your phone while in the cleanroom, double glove, and change your gloves often.

- For the following fabrication processes detailed in this appendix, glassware should be organized in sets: no metals, metals, gold (if gold is used for bond pads). Each set should contain a dish for DI water, developer, solvents, and acids. It’s a good idea to have individual plastic beakers for acetone and IPA as well for each set. Certain cleanroom tools are sensitive to metals contamination like the furnaces, so it is important to keep track of which dishes have been exposed to metals to avoid transferring metal particles to your sample and then
to a sensitive tool. Gold is an especially egregious contaminant for other silicon devices that must be kept track of.

- For RCA cleaning, I use a dedicated set of fused quartz glassware. As it turns out, normal borosilicate glassware is composed of ∼5% sodium oxide which can supposedly leach out of the glassware at elevated temperature and contaminate your samples. I don’t know the amount of sodium that will leach out of borosilicate glass but it’s better to be safe. Only a part in 10 million of sodium in surface concentration is needed to kill a device’s mobility. Fused quartz glassware is very high purity SiO$_2$ and as such will not transfer sodium to your sample.

- Check the surface of your sample under the optical microscope after each cleaning step to ensure the no residue is present on the surface. A clean sample should not have anything visible on the surface.

- New glassware should be cleaned with Piranha solution prior to being used. Glassware should be cleaned after each use with liberal amounts of IPA (for developer and solvent dishes) or DI water (for acid and DI water dishes) and blow dried. Clean glassware should never have a visible residue.

- Each glassware set should also have dedicated tweezers: carbon fiber tipped stainless steel tweezers for general use and plastic (delrin) tweezers for acids. These two types of tweezers will cover most general uses without being attacked by cleanroom processes. The carbon fiber tipped tweezers are the absolute best in terms of handling small chips. They avoid the danger of chipping the edges of the chip (like metal tweezers tend to) and also avoid the danger of dropping or flinging your chip (like some of the flimsier plastic tweezers tend to). Some cleanroom processes require that you use tweezers of specific materials (the
Savannah ALD requires metal tweezers and the oxidation furnaces forbid metal tweezers), so it is also a good idea to have dedicated tweezers for those tools.

- During solvent cleaning steps, before removing your sample from the IPA to blow dry, it is important to rinse your tweezers with IPA and to blow dry them before using them to remove your sample. It is very easy to transfer residual photoresist containing solvent from your tweezers back to your sample otherwise.

- In a similar vein, samples should never be allowed to air dry. Contaminants in the solvent will re-adhere to the surface of the sample and will be very difficult to remove afterwards. Samples should always be rinsed liberally in IPA at the last step of solvent cleaning so that the residual contaminant concentration is low. Then the sample should be quickly blow dried in $\text{N}_2$. IPA will cleanly and smoothly be blown off the surface of a clean sample.

- Never use the city water taps instead of the DI water taps.

- Any time the silicon dioxide layer is exposed in the MOSFET it is critical to keep the sample clean. Avoid doing wet processes as much as possible until the oxide encapsulated with the gate material.

- In general, never use your cleanroom tools or glassware outside of the cleanroom. Wipe down the outside of your cleanroom box with IPA often as well. Wear gloves when carrying your cleanroom box in and out of the cleanroom to avoid transferring finger grease to the handle of your box. Finger grease present on the handle of the box will be transferred to your tools and glassware inside of the cleanroom.
A.2 MOSFET Fabrication

This section details the fabrication process for n-type inversion silicon MOSFETs yielding mobilities of 23,000 cm$^2$/Vs. Hall bars and large area MOSFETs for ESR measurements are both fabricated from this process and differ simply in their gate geometries. These MOSFETs are fabricated from commercially produced gate stacks. Common failure modes and troubleshooting tricks will be detailed at the end.

A.2.1 Detailed Fabrication Process

1. Substrate (Novati Technologies): High-resistivity (1,000-3,000 Ω-cm), p-type (boron), (100) orientation, 8 inch wafers. 30 nm dry, thermal oxide grown with Dichloroethane. 200 nm low-pressure chemical vapor deposition (LPCVD) amorphous silicon.

2. Cleave wafer into 3 inch squares for further processing. Solvent clean: 3x rinse in Acetone, 3x rinse in IPA. Blow dry in N$_2$.

3. Photolithography with AZ1505 photoresist to define contact holes for source and drain contacts.

4. SAMCO800 etch to etch contact holes through the a-Si cap. 40 cycles, recipe 4.

5. Tepla O$_2$ 1 minute descum to etch to etch fluoro polymer deposited by SAMCO800.


7. SiO$_2$ etch - BOE 10:1 1 min. 3x in DI water.
8. Send out wafer for ion implantation - blanket Arsenic implant to dope source, drain and gate. 35 keV Arsenic, $5 \times 10^{15} \text{cm}^{-2}$, Leonard Kroko Inc. This implantation recipe gives a self aligned source/drain.


10. Dice wafer into dies for individual processing.

11. Photolithography with AZ1505 photoresist to define gate geometry.

12. SAMCO800 etch to define gate geometry through the a-Si cap. 40 cycles, recipe 4.

13. Tepla O$_2$ 1 minute descum to etch to etch fluoro polymer deposited by SAMCO800.


15. RCA clean 1, DI water rinse 3x, RCA clean 2, DI water rinse 3x.

16. High temperature anneal to activate dopants, crystallize the a-Si and anneal out damage from the ion implant - 900 C, 1 hour in Tystar oxidation tube in N$_2$. The Tystar is no longer commissioned so the option now is to use the CVD annealing furnace (tube 3) or the CVD oxidation furnace (tube 1).

17. Photolithography with AZ nLOF2070 to deposit aluminum contacts on source, drain and gate.

18. HF dip to etch off native oxide from contacts - 1 minute 100:1 HF, 3x rinse in DI water. Silicon surfaces should be hydrophobic after this step. Quickly transfer sample into a vacuum box and proceed to next step.

20. O\textsubscript{2} strip 2 minutes in Tepla to etch off residual photoresist

21. Forming gas anneal - 25 minutes in Thermco metals annealing tube (tube 2),
   or CVD metals annealing tube (tube 2). 435 C, 5% H\textsubscript{2}, 95% N\textsubscript{2}.

22. Photolithography with AZ nLOF 2070 to deposit gold solder pads.

23. Evaporate 20 nm titanium in Lyon Edwards metals evaporator. Without breaking vacuum, evaporate 100 nm gold. Liftoff in TechniStrip NI555 at 80 C for 1 hour at least. Solvent clean.

24. Dice die into individual devices.

A.2.2 Notes on the Geometry and Placement of the Source/Drain Contacts in Large-Area MOSFETs

The specific geometry of our large-area MOSFET is an elongated channel where the vertical edge is 2 cm long and the horizontal edge is 0.3 cm wide in order to fit within a volume microwave resonator. The source/drain contacts are located at the top and bottom of the channel. The placement of the source/drain contacts went through several iterations before settling on this design. Originally both the source and drain were placed at the top of the device with leads extending downwards to the gate (Figure A.1). The reason for this design was to ensure that the degenerately doped source and drain regions are located outside of the microwave resonator as these regions will contribute a background ESR signal near the g-factor of 2DEG signal. This background signal is more prominent for phosphorus than arsenic implants and sharper at lower densities. We found that with this device design, at temperatures lower than 1 K, the long narrow leads from the gate to the source/drain actually impede electrons trying to escape to the source and drain. We then experimented with the source drain placed directly on the edges of the gate along the vertical edges.
and horizontal edges. As it turns out, having the source and drain in the resonator does not pose an issue because our arsenic implant dose is high enough that the background signal is immeasurable. With the source/drain on the vertical edges of the gate, we found that the ESR signal become extremely noisy when measuring the 2DEG signal near the threshold voltage of the device. This is because transverse electric fields within the resonator can induce currents in the 2DEG which add noise to the ESR signal \cite{88}. Having the source and drain on the horizontal edges of the gate circumvents all of the above issues.

### A.2.3 Common Failure Modes and Troubleshooting

This section details common failure device failure modes and troubleshooting methods.
1. Gate leakage: The biggest failure mode of these devices is leakage from the 2DEG to the gate. Gate leakage can be tested at room temperature with the semiconductor parameter analyzer.

Avoid using digital multimeters (DMM) on the resistance setting to measure the gate to 2DEG resistance. On the auto ranging setting, the DMM will try to push a known current through the terminals and will measure the resultant voltage drop across the terminals to calculate the resistance. It will keep pushing the voltage across the terminals to achieve this current up to 50 V which will cause dielectric breakdown of a good (non-leaky) oxide.

Leakage currents of < 1 µA at 1 V at room temperature will generally freeze out at helium temperatures resulting in leakage currents of <1 nA, and typically on the order of 10’s of pA. This is an acceptable level of gate leakage. Gate leakages much larger than 1 µA at 1 V will not freeze out and are indicative of a short from the gate to the 2DEG.

Always hand solder wires to the bond pads with the least amount of pressure possible to avoid punching through the oxide. An alternative, but less tested method, is to use the West ball bonder which I have used for only a few devices but have had good luck with. Avoid using the Questar auto-bonder as it will punch through 30 nm of oxide and create a short from the gate to the 2DEG.

In the same vein, when mounting wafers to be diced in the dicing saw, I usually do not use the tape roller to mount the wafer to the tape. Instead I will apply the tape to the metal o-ring and cut it out, then use wafer tweezers to lay the wafer on top of the sticky side of the tape and work out any bubbles between the tape and the wafer manually with a glove on.
Hall bars and large area MOSFETs are relatively robust to electrostatic discharge from their large gate areas (>1 mm²), but it is a good idea to ground yourself while soldering the devices.

2. Contact resistance: In early versions of these devices, we used a very shallow ion implantation recipe (5 keV, 5×10¹⁵ cm⁻² Phosphorus) because we were afraid of the ion implantation penetrating through the a-Si gate and damaging the underlying oxide. It turns out that the implanted source/drain regions of these devices did not diffuse very far underneath the gate oxide and the devices had significant amount of contact resistance. These devices required ~100 mV across the source drain to induce current through the device. The drain current around 0 source drain voltage should be ohmic, e.g. perfectly linear through 0 voltage.

This issue has been solved with the current implantation recipe (35 keV, 5×10¹⁵ cm⁻² Arsenic). If contact resistance is observed with this recipe, it probably means the oxide has been undercut in the region where the gate overlaps the source/drain implant. This can happen during the BOE etch of the SiO₂ if the etch goes on for too long.

A.3 Quantum Dot Fabrication

A.3.1 Detailed Fabrication Process

1. Substrate (Novati Technologies): High-resistivity (1,000-3,000 Ω-cm), p-type (boron), (100) orientation, 8 inch wafers. 30 nm dry, thermal oxide grown with dichloroethane. 200 nm low-pressure chemical vapor deposition (LPCVD) amorphous silicon
2. Cleave wafer into 3 inch squares for further processing. Solvent clean: 3x rinse in Acetone, 3x rinse in IPA. Blow dry in N₂.

3. Photolithography with AZ1505 photoresist to define contact holes for source and drain contacts.

4. SAMCO800 etch to etch contact holes through the a-Si cap. 40 cycles, recipe 4. Strip resist.

5. SiO₂ etch - BOE 10:1 1 min. 3x in DI water.

6. Photolithography with AZ1518 photoresist to define 300 µm ion implantation windows.

7. Send out wafer for ion implantation - blanket Arsenic implant to dope source, drain and gate. 35 keV Arsenic, 5×10¹⁵ cm⁻², Leonard Kroko Inc. This implantation recipe gives a self aligned source/drain.

8. Strip resist overnight in 1165 heated to 80 C, then solvent clean and O₂ strip in Tepla if necessary.

9. RCA clean and anneal at 900 C to activate implants.

10. Dice wafer into 6 mm chips for individual device processing.

11. E-beam lithography layer 0 (PMMA 950K): define registration markers. Evaporate 150 nm of Al from Lyon Al evaporator, lift-off overnight in 1165 heated to 80 C, solvent clean.

12. Photolithography with AZ1505 to thin down implanted polysilicon gate area.


15. SAMCO 800 etch to etch polysilicon depletion gates (15 cycles, recipe 4). Strip resist with Tepla. Oxide is exposed in this step, so be careful with wet processes.

16. As soon as resist is stripped, immediately deposit 200 cycles of ALD aluminum oxide (Savannah ALD).

17. Photolithography with AZ1505 to etch ALD to open holes to the degenerately doped contacts.

18. 1 min 30 second 10:1 BOE etch. Strip resist.

19. E-beam lithography layer 3 (PMMA 950K): plunger gates and reservoir gates. Evaporate 150 nm of Al from Lyon Al evaporator, lift-off overnight in 1165 heated to 80 C, solvent clean.

20. 10 minute O$_2$ strip in Tepla to oxidize aluminum gates.


22. Photolithography to define Al bond pads. Use AZ nLOF to do liftoff. Evaporate 400 nm Al, liftoff, solvent clean.

23. Photolithography with AZ 1505 to etch individual Al bond pads. Etch Al in Al etchant for about 10 minutes or until Al is visibly etched. Strip resist.

24. Forming gas anneal in CVD 2 (metals annealing) furnace. 30 minutes, 400 C, forming gas.
25. Mount sample to PCB with Elmer’s rubber cement, wire-bond to sample with Questar auto-bonder.

A.4 Recipes

A.4.1 Photolithography

Photolithography with AZ15xx Photoresist

AZ15xx is a family of positive tone resists sensitive to 310-440 nm light (i, h, g-line mercury light). The ‘xx’ denotes the thickness of the resist in hundreds of nanometers of the resist film spun at 4000 rpm

- Bake sample at 95°C for 1 minute to drive off residual solvents.
- Spin on HMDS at 4000 rpm for 40 seconds. HMDS improves photoresist adhesion to silicon surfaces.
- Spin on AZ15xx photoresist at 4000 rpm for 40 seconds.
- Soft bake sample at 95°C for 1 minute to evaporate photoresist solvents.
- Expose sample for 10 seconds in the MJB4, or do direct write with the Heidelberg DWL 66+. Refer to the Heidelberg manual for the most up to date exposure parameters.
- Develop sample in AZ300MIF for 1 minute. Gently swirl the beaker while the sample develops.
- Rinse sample in DI water 3x.
- Blow dry sample with N₂.
Photolithography with AZ nLOF2070 Photoresist

AZ nLOF is a negative tone lift-off resist sensitive to 365 nm light (i-line). The top layer of resist absorbs the majority of the exposure dose and crosslinks after a post exposure bake, leaving bottom layer of resist soluble. This creates an undercut in the resist profile after development which allows for lift off. I have diluted nLOF2070 with EBRPG (Edge bead remover propylene glycol) in a 1:1 volume ratio to get \( \sim 2 \mu m \) resist thickness.

- Bake sample at 95 C for 1 minute to drive off residual solvents.
- Spin on HMDS at 4000 rpm for 40 seconds. HMDS improves photoresist adhesion to silicon surfaces.
- Spin on nLOF photoresist at 4000 rpm for 40 seconds.
- Soft bake sample at 110 C for 1 minute to evaporate photoresist solvents.
- Expose sample for 10 seconds in the MJB4. Do not use the Heidelberg for direct write as it utilizes a 405 nm laser which nLOF is not sensitive to.
- Post exposure bake at 110 C for 1 minute.
- Develop sample in AZ300MIF for 2 minutes. Gently swirl the beaker while the sample develops.
- Rinse sample in DI water 3x.
- Blow dry sample with \( N_2 \).
- Evaporate metal onto sample.
- Use TechniStrip NI555 (oxalic acid) at 80 C to lift off metal. Full lift off usually takes about an hour. An IPA spray bottle can be used to help lift off the
metal. If metal remains after an hour, transfer the sample into a fresh beaker of TechniStrip and sonicate for 10 seconds at a time until metal has lifted off.

- Solvent clean the sample afterwards and blow dry.

**A.4.2 Electron-beam Lithography**

These e-beam recipes are for the Elionix F125 E-beam system in Jadwin.

**E-beam lithography with PMMA950K A4**

- Bake sample at 180 °C for 1 minute to drive off residual solvents.
- Spin on PMMA at 4000 rpm for 40 seconds.
- Bake at 180 °C for 15 minutes.
- Exposure parameters: 100 μm write field, 1 nA beam current, 1,600 μC/cm².
- Develop in 1:3 MIBK:IPA solution for 1 minute.
- Rinse in IPA.

**E-beam lithography with ZEP520a 1:1**

- Bake sample at 180 °C for 1 minute to drive off residual solvents.
- Spin on PMMA at 4000 rpm for 40 seconds.
- Bake at 180 °C for 15 minutes.
- Exposure parameters: 100 μm write field, 1 nA beam current, 360 μC/cm².
- Develop in ZED-N50 solution for 1 minute.
- Rinse in IPA.
- ZEP520a may be stripped using ZDMAC.
A.4.3 Solvent Clean

- Soak in acetone for 10 minutes. Rinse 3 times in fresh acetone.
- Soak in isopropanol (IPA) for 10 minutes. Rinse 3 times in fresh IPA.
- Rinse tweezers in IPA and blow dry.
- Remove sample from IPA and blow dry immediately with $N_2$.

A.4.4 RCA Clean

- RCA 1: 5 parts DI water, 1 part ammonium hydroxide, 1 part hydrogen peroxide. Add DI water to fused quartz beaker, heat to 80 C. Add ammonium hydroxide then hydrogen peroxide.
- RCA 2: 5 parts DI water, 1 part hydrochloric acid, 1 part hydrogen peroxide. Add DI water to beaker, heat to 80 C. Add hydrochloric acid then hydrogen peroxide.
- Immerse sample in RCA 1 mixture for 15 minutes at 80 C. This step will remove organic contaminants and certain metallic contaminants.
- Remove sample from mixture and immerse immediately into DI water. Rinse sample in DI water 5x using the overflow method if possible. Transfer sample directly from DI water to RCA 2 mixture.
- Immerse sample in RCA 2 mixture for 15 minutes at 80 C. This step will remove various metallic contaminants.
- Remove sample from mixture and immerse immediately into fresh DI water. Rinse sample in DI water 5x using the overflow method if possible.
- Remove sample from mixture and immediately blow dry with $N_2$. 
• Note 1: It is a good idea to clean your tweezers in Piranha solution while the sample is immersed in RCA 1. Use delrin tweezers. Rinse tweezers well with DI water and blow dry between each transfer step.

• Note 2: Some people do also do an HF dip as part of the RCA clean procedure. If this is done, care must be taken to avoid re-adhering contaminants on the surface of the solution to the surface of the sample. The bare silicon surface is extremely reactive and contaminants will stick to the surface.

### A.4.5 Piranha Clean

• 4 parts sulfuric acid, 1 part hydrogen peroxide. Add the sulfuric acid to the beaker first then the hydrogen peroxide.

• Immerse sample in mixture for 15 minutes.

• Remove sample with delrin (or teflon) tweezers and transfer to DI water. Piranha will attack carbon fiber tipped tweezers and metal. Rinse sample in DI water 5x.

• Caution: Piranha solution will get very hot as it reacts. It will also etch straight through Tex wipes and clothing.

### A.4.6 Resist Stripping

For general photoresist stripping, PRS1000 stripper is good enough followed by a standard solvent clean. If, however, the resist has been carbonized from a reactive ion etch or ion implantation, more aggressive stripping techniques maybe needed.

• Soak sample in PRS1000 stripper for 10 minutes.

• Solvent clean.
Optional steps:

• Heat PRS1000 to 80 C.

• Sonicate.

Notes:

• 1165 is a more aggressive stripper than PRS1000 and may be used if PRS1000 is not sufficient. Note that when heated, 1165 will attack plastic tweezers.

• The Tepla O$_2$ strip recipes are good for safely removing carbonized photoresist residue without attacking other materials. It will oxidize a few nanometers of exposed silicon.

• A Piranha clean may also be used to remove persistent organic residues. Piranha will also attack metals, however.

A.4.7 Wet Chemical Etching of Polysilicon

There are several methods of wet etching silicon for example KOH solution, ethylenediaminepyrocatechol and mixtures of HF and Nitric acids. KOH will contaminate silicon dioxide and ethylenediaminepyrocatechol will not etch n-type degenerately doped silicon. We therefore use a mixture of HF and Nitric acid diluted with DI water. Nitric acid is a strong oxidizer which will attack photoresist after some time. The recipe developed here was adopted from https://www.seas.upenn.edu/nanosop/documents/IsotropicSiliconEtch.pdf and will etch about 1 $\mu$m per minute and will leave a rough surface and rough feature edges.

• Do standard photolithography with AZ1518 photoresist.

• Hard bake the sample for 40 minutes at 120 C.

• In a polyethylene beaker mix 49% HF and Nitric acid in a 1:1 ratio.
• In a separate polyethylene beaker add 20 ml of DI water.

• Add 30 ml of the HF/Nitric acid mixture to the DI water beaker.

• Prepare a third polyethylene beaker and fill with DI water to quench the reaction.

• With teflon tweezers, add the sample to the acid/DI water mixture.

• Etch the sample for 15-20 seconds.

• Transfer the sample to the DI water beaker and rinse 3x in DI water.

Notes: I’ve found that more dilute mixtures of this acid etch will etch more slowly and controllably but will tend to undercut the photoresist, especially in the regions where the gate is self aligned to the n+ contacts. In devices where this happens, the device will have significant contact resistance at helium temperature.

A.4.8 Savannah ALD Aluminum Oxide Growth Conditions

Use metal tweezers for this process. ALD growth is nominally 1.1 angstrom/cycle.

• N\textsubscript{2} flow rate: 20 sccm.

• 150 C inner and outer rings.

5 cycles water to prime surface of sample

<table>
<thead>
<tr>
<th>Valve</th>
<th>Pulse (s)</th>
<th>Expose (s)</th>
<th>Pump (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 (water)</td>
<td>0.1</td>
<td>0</td>
<td>15</td>
</tr>
<tr>
<td>0 (water)</td>
<td>0.1</td>
<td>0</td>
<td>15</td>
</tr>
</tbody>
</table>

200 cycles water and tri-methyl aluminum (TMA)
### Valves and Exposure Times

<table>
<thead>
<tr>
<th>Valve</th>
<th>Pulse (s)</th>
<th>Expose (s)</th>
<th>Pump (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 (water)</td>
<td>0.1</td>
<td>0</td>
<td>15</td>
</tr>
<tr>
<td>1 (TMA)</td>
<td>0.1</td>
<td>0</td>
<td>15</td>
</tr>
</tbody>
</table>

#### A.4.9 Chlorine Cleaning of Furnace Tubes

If a furnace tube has been contaminated with sodium, it can be cleaned with a chlorine containing ambient such as dichloroethylene, sold commercially as Trans-LC. Below is a rough outline of how to clean the furnace tube according to former PRISM staff member Conrad Silvestre:

- Heat furnace tube to $\sim 1000$ C
- Bubble Trans-LC through the tube for 1-2 hours to clean sodium and other alkali contaminants
- Flow mixture of $O_2$ and $N_2$ through the tube for 1-2 hours to combust any organic contaminants and residual carbon from the Trans-LC
- Flow $N_2$ through the tube for 1-2 hours to flush out any residual particulates
- Cool furnace back to its normal standby temperature
Bibliography


147


