Molecular and Quantum Dot Floating Gate Non-Volatile Memories

by

Hassen Abdu

Submitted to the Department of Electrical Engineering and Computer Science in partial fulfillment of the requirements for the degree of Master of Engineering in Electrical Engineering and Computer Science at the MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2008

© Massachusetts Institute of Technology 2008. All rights reserved.

Author .................................................................

Department of Electrical Engineering and Computer Science

May 15, 2008

Certified by ..............................................................

Vladimir Bulovic
Professor

Thesis Supervisor

Accepted by ...............................................................

Arthur C. Smith
Chairman, Department Committee on Graduate Students
Molecular and Quantum Dot Floating Gate Non-Volatile Memories

by

Hassen Abdu

Submitted to the Department of Electrical Engineering and Computer Science on May 15, 2008, in partial fulfillment of the requirements for the degree of Master of Engineering in Electrical Engineering and Computer Science

Abstract

Conventional Flash memory devices face a scaling issue that will impede memory scaling beyond the 50nm node: a reliability issue involving the tunneling oxide thickness and charge retention. A possible solution is to replace the continuous floating gate, where charge is stored, with a segmented charge storage film, so that leakage through defects in the tunneling oxide would be localized. We first explored using quantum dots as possible floating gate replacements. After conducting simulations, we established the need for the smallest possible segmented structures. This led us to the use of molecular films as floating gates in non-volatile flash memories. As an example, a single organic molecule of 3,4,9,10-parylene tetracarboxylic dianhydride (PTCDA) occupies 1nm² in area and is capable of storing and retaining a single charge. If a defect is present in the tunneling oxide below the floating gate, only a few molecules of PTCDA would be affected due to poor lateral conduction between PTCDA molecules. We can, therefore, project that such molecular thin films of PTCDA are likely to meet demanding size and packing density requirements of advancing flash memory technology.

Thesis Supervisor: Vladimir Bulovic
Title: Professor
Acknowledgments

I would like to thank God for blessing me with this opportunity to study and learn with great people at such a fine institution like MIT. I pray that I can make a positive impact on the world with the knowledge I have acquired here.

I would like to thank my family, especially my mother and my sister for being supportive and understanding as I pursued an education so far away from home for such a long time. I love them dearly and I will always be a devoted son and brother to them.

I would like to thank everyone with whom I have formed strong friendships. I hope I was as much of a committed friend to you as you have been to me.

I would like to thank my research colleagues who have taught me so much. Specifically, I would like to thank my advisor, Professor Vladimir Bulovic, for being an encouraging mentor. Also, I would like to thank Ivan Nausieda, Polina Anikeeva, Osama Nayfeh, Joshua Leu, and everyone else in the Laboratory of Organic Optics and Electronics group. I have gained so much respect for all of you and I am confident that I will see many of you accomplishing great things in the future.

I would like to thank you, for deciding to read my thesis. I hope you can learn and benefit from my work on molecular and quantum dot floating gate memories.
# Contents

1 Flash Memory ...................................................... 15
  1.1 Introduction to Flash Memory ................................. 15
    1.1.1 Memory Industry and Applications ........................ 15
    1.1.2 Structure of Flash Memory ................................. 17
    1.1.3 Charging Mechanisms: Programming/Erasing ............... 18
  1.2 Limitations of Flash Memory .................................. 20
  1.3 Future of Flash Memory ....................................... 23

2 Metal-Oxide-Semiconductor (MOS) Capacitors ................ 25
  2.1 Introduction to MOS Capacitors ............................... 25
    2.1.1 MOS Capacitor in Thermal Equilibrium ...................... 25
    2.1.2 MOS Capacitor under Applied Bias .......................... 27
  2.2 Experimental Setup .......................................... 32

3 Quantum Dot Floating Gate Non-Volatile Memory ................. 33
  3.1 Introduction to Quantum Dot Floating Gate Non-Volatile Memory .... 33
  3.2 Introduction to Quantum Dots ................................ 34
  3.3 Fabrication of QD Floating Gate Capacitor .................... 35
    3.3.1 RCA Clean ..................................................... 35
    3.3.2 Oxidation ..................................................... 36
    3.3.3 Spin-Casting Quantum Dots .................................. 37
    3.3.4 Parylene Deposition ....................................... 38
    3.3.5 Gold Deposition ............................................. 39
### 3.4 Experimental Results of QD Capacitor

3.5 A Perspective on the Long-term Viability of Quantum Dot Floating Gate Memories
- 3.5.1 Simulations
- 3.5.2 Conclusion

### 4 Molecular Floating Gate Memory

4.1 Introduction to Molecular Floating Gate Non-volatile Memory
4.2 Introduction to Organic Molecular Materials
4.3 Fabrication of Molecular Floating Gate Capacitor
- 4.3.1 Evaporation of PTCDA
- 4.3.2 Plasma Enhanced Chemical Vapor Deposition (PECVD) of SiO₂
4.4 Experimental Results of Molecular Floating Gate Capacitor
4.5 Conclusion

### 5 Future Work

5.1 Protective Layer

### 6 Conclusion

### A Quantum Dot Size Simulation Code
List of Figures

1-1 Memory Market Trend [21] ................................................. 16
1-2 Non-Volatile Memory Market [21] ........................................ 16
1-3 Flash Memory Cell ............................................................. 17
1-4 Flash Memory Cell Array ..................................................... 18
1-5 IV Curves for a Floating Gate Cell [22] .............................. 18
1-6 Flash Memory Energy Band Diagram [22] ............................ 19
1-7 Fowler-Nordheim tunneling current as a function of Electric Field [1] 20
1-8 Maximum Floating-gate Height to Limit Floating Gate Capacitance Coupling Interference [14] ................................................. 21
1-9 Example Endurance test of a Memory Cell: Decreasing Threshold voltage window with Increasing Cycling [3] .......................... 22
1-10 Number of Electrons Stored in a Floating Gate and Tolerated Charge Loss [14] .......................................................... 22
1-11 Roadmap for NAND Flash Memories [14] .......................... 23

2-1 Structure of a MOS Device [13] ........................................... 26
2-2 Energy Band Diagram for a MOS capacitor [22] .................... 27
2-3 CV for P-type Semiconductor Substrate [13] .......................... 28
2-4 MOS Operation Modes [13] ................................................. 29
2-5 MOS Energy Band Diagram [22] ........................................... 29
2-6 MOS Capacitor CV with Frequency Dependence. a) low frequency, b)Intermediate frequency, c) high frequency, d)high frequency with fast sweep [22] ..................................................... 31
3-1 Quantum Dot Memory ........................................... 34
3-2 Device Structure of Quantum Dot Floating Gate Capacitor .... 35
3-3 Quantum Dot Deposition via Phase Segregation [5] ............. 37
3-4 Atomic Force Microscope Image of Quantum Dot Calibration (1:7 QD:Chloroform) ............................................. 37
3-5 Chemical Structure of Parylene Varients .......................... 38
3-6 Shadow Mask for Top Metal Contact (Diameters: A=1mm, B=0.5mm, C=0.2mm) .................................................. 39
3-7 CV Measurement of QD Device ..................................... 41
3-8 QD Retention Measurements ........................................ 43
3-9 Quantum Dot Device Retention Measurement .................... 44
3-10 (left) Unordered Monolayer of QDs. (right) Hexagonally Closed-Packed Ordered monolayer of QDs. The QDs are part of the device if encompassed by the top gate contact, which is represented by the square. ........ 45
3-11 Scanning Electron Microscope (SEM) image of Quantum Dot array. Red region indicates QD locations detected by software. Simulated analysis is based on detected locations from SEM images. .......... 46
3-12 Random Monolayer of QDs with Translation and Rotation. Square side length-QD size ratio (R) of 8 and 14 are needed to obtain variations (V) of +/-10% and +/-5%, respectively. .................. 47
3-13 Scanning Electron Microscope (SEM) image of ordered Quantum Dot array. QD locations detected by software. Simulated analysis is based on detected locations from SEM images. ............... 47
3-14 Ordered Monolayer of QDs with Translation and Rotation variability. Square side length-QD size ratio (R) of 8 and 14 are needed to obtain variations (V) of +/-10% and +/-5%, respectively. ............... 48
3-15 Ordered Monolayer of QDs with Translation along a grain axis and Rotation. Square side length-QD size ratio (R) of 7 and 13 are needed to obtain variations (V) of +/-10% and +/-5%, respectively. .......... 49
4-1 Molecular Floating Gate Memory Transistor ........................................... 51
4-2 Molecular Structure of 3,4,9,10-perylenetetracarboxylic dianhydride (PTCDA) [10] ................................................................. 52
4-3 Crystal Structure of 3,4,9,10-perylenetetracarboxylic dianhydride (PTCDA) [10] ................................................................. 53
4-4 Scanning Tunneling Microscope of 7 x 7 nm² Monolayer of PTCDA on Ag [8] ............................................................................. 53
4-5 Device Structure of a PTCDA Capacitor .................................................. 54
4-6 Carrier Mobility in PTCDA as a Function of Deposition Rate [10] ........ 55
4-7 Atomic Force Microscope Image of 50nm PTCDA ................................. 56
4-8 Capacitance vs. Voltage for a PTCDA Floating Gate Capacitor Device at 100kHz frequency ..................................................... 58
4-9 Energy Band Diagram for a PTCDA Floating Gate Capacitor ............. 59
4-10 Endurance Test ................................................................................... 60
4-11 CV curves ........................................................................................... 61
5-1 AFM of 5nm Al .................................................................................... 64
5-2 AFM of 10nm Al .................................................................................. 65
List of Tables

3.1 Oxidation Constants [15] .......................... 36
3.2 QD-to-QD Distance for Technology Design Nodes ............. 49
Chapter 1

Flash Memory

1.1 Introduction to Flash Memory

There are two main types of memory: volatile memory, which requires that power is supplied to retain data, and non-volatile memory, which retains data even when power has been disconnected. Flash memory, a popular type of non-volatile memory, gets its name from its ability to write a cell individually and erase a large amount of cells by electrically "flashing" them at the same time. Flash memory has become ubiquitous and has become a crucial part of our day to day lives as they are found in our cell phones, cameras, personal computers, and many other electronic devices. However, after years of intense growth in the Flash memory market, Flash memory faces scaling limitations that impedes its future.

1.1.1 Memory Industry and Applications

The semiconductor memory industry is currently about a $60B a year industry, Figure 1-1, tripling in size over the past 10 years [3]. The future of the memory industry will be very lucrative, especially as the demand for Flash memory continues to grow at a fast pace. As shown in Figure 1-1, Flash memory has increased 5-fold in 10 years, which is about 18% annual growth. Also, as shown in Figure 1-1, non-volatile memory currently accounts for nearly 50% of the memory market, and it is expected
to grow to over $60B, as shown in Figure 1-2.

![Figure 1-1: Memory Market Trend [21]](image)

Non-volatile memory is further divided into three categories: Flash, EPROM (electrically programmable read only memory), and EEPROM (electrically erasable programmable read only memory). EPROM, the first of non-volatile memory from the 1980s, was difficult to use in integrated systems because ultraviolet exposure was required to erase the memory. EEPROM was invented to overcome this inconvenience, however, due to its erasing approach by a single byte, these memory devices are much larger, which in the semiconductor industry means it is more expensive [19]. Flash memory became very popular in the mid-1990s as a nice compromise between cost and flexibility, only after reliability issues were resolved.

![Figure 1-2: Non-Volatile Memory Market Market [21]](image)

There are two architectures for Flash memory: NOR Flash and NAND Flash. They are different in that NOR Flash is optimized for executing computer programs...
in place by having faster random-access reading times. While NAND Flash is optimized for data storage by having faster write times. Optimization for each Flash memory architecture is achieved by choosing between parallel and serial reading and programming, and choosing which charging mechanism is used for programming and erasing, as discussed below. As shown in Figure 1-2, NAND memory is expected to become much more popular than NOR in 2011, taking about 70% of the Flash memory market.

1.1.2 Structure of Flash Memory

A Flash memory cell is a metal-oxide-semiconductor (MOS) transistor with a continuous, conductive floating gate (FG) sandwiched between a tunneling dielectric and a gate dielectric, as shown in Figure 1-3. These cells are arranged by connecting the terminals of each cell to each other via bit lines, word lines, and source lines, as shown in Figure 1-4. By applying voltages to the word and bit lines, we can read, erase, and program individual Flash cells.

![Flash Memory Cell](image)

Figure 1-3: Flash Memory Cell

The Flash cell operates by charging and discharging charge carriers in the floating gate. By changing the amount of charge in the floating gate, the threshold voltage ($V_T$) changes by

$$V_T = V_{T0} - Q/C_{ox}$$

where $V_{T0}$ is the initial threshold voltage, $Q$ is the charge, and $C_{ox}$ is the effective oxide capacitance (a series combination of the gate dielectric capacitance and tunneling dielectric capacitance). This threshold shift is proportional to the amount of charge
stored and consistency across memory cells is crucial to overall device operation.

This shift in threshold voltage affects the amount of current that flows when applying a constant voltage to the gate, a voltage we apply when reading the cell. Therefore, to create logic states “1” and “0”, we decide that if we measure a high current when the reading voltage is applied that state will equal a “1” and if we measure a low current that state will equal a “0”. As an example, refer to Figure 1-5.

Figure 1-5: IV Curves for a Floating Gate Cell [22]

1.1.3 Charging Mechanisms: Programming/Erasing

Charge carriers need to get from the channel between the source and drain (refer to Figure 1-6) to the floating gate. However, there is a large potential barrier in the form
of a tunneling dielectric. There are two main mechanisms by which charge carriers can charge and discharge the floating gate: Hot-Electron Injection and Fowler-Nordheim tunneling.

Figure 1-6: Flash Memory Energy Band Diagram [22]

Hot-Electron Injection is a phenomenon by which a charge carrier in the channel gains energy from the electric field that exists laterally between the source and drain, and then crosses the oxide energy barrier into the floating gate by experiencing a vertical electric field between the control gate and substrate. There are several quantum mechanical models to quantify this mechanism, but it is beyond the scope of this thesis.

Fowler-Nordheim Tunneling is another mechanism for charge carriers to cross the oxide energy barrier. The concept of tunneling is rooted in quantum mechanics. We are saying that a particle can penetrate a forbidden region in order to tunnel from one classically allowed region (substrate) to another (floating gate). However, this phenomenon is probabilistic, depending on the source material, and the height and width of the oxide barrier. Also, the expression for current density is derived from a free-electron gas model and WKB (Wentzel-Kramers-Brillouim) approximation:

$$J = \frac{q^3 F^2}{16\pi^2 \hbar^2 \phi_B} e^{-\frac{1}{2} \left(\frac{2m^*_e}{\hbar^2}\right)^{1/2} \phi_B^{3/2}}$$

where $\phi_B$ is the barrier height, $m^*_e$ is the effective mass of the electron, $\hbar$ is Planck’s constant, $q$ is the electron charge, and $F$ is the electric field through the oxide [19].

A plot of the amount of Fowler-Nordheim tunneling as a function of electric field is shown in Figure 1-7. Though we can increase Fowler-Nordheim tunneling current by increasing voltage or decreasing the oxide thickness, we must balance performance
and reliability needs. The oxide cannot be made too thick as it would significantly increase the voltage and time required for charging the floating gate nor can it be made too thin as it would increase oxide defect density and ruin device reliability. The issue of device reliability is explained further in Section 1.2.

Figure 1-7: Fowler-Nordheim tunneling current as a function of Electric Field [1]

1.2 Limitations of Flash Memory

Non-volatile Flash memory has achieved widespread popularity and success due to its reliable performance. Memory devices can retain data for a minimum of 10 years and can withstand over $10^6$ program/erase cycles [2]. As memory devices continue to gain even greater integration into mobile, electronic devices, reliability will continue to be paramount. However, reliability is becoming very hard to maintain as we scale down Flash memory.

There are many limitations facing Flash memory in the future. One is floating gate capacitance coupling between memory cells [20]. As the distance between memory cells decreases, the charge on one floating gate can cause a significant threshold voltage shift on adjacent memory cells. This problem can be resolved by scaling down the vertical dimension of the floating gate, as shown in Figure 1-8. There are other physical and electrical challenges to scaling Flash memory, for instance short channel
effects [14], however the major challenge of focus in this paper is reliability.

![Figure 1-8: Maximum Floating-gate Height to Limit Floating Gate Capacitance Coupling Interference [14]](image)

Reliability of a Flash cell is attributed to its endurance and retention. As mentioned before, a floating gate needs to endure over $10^5$ program/erase cycles. An endurance test is a good way to measure the degradation of the tunneling oxide over many cycles - a test that will replicate the kind of usage a memory cell will handle. The endurance test is conducted by applying program and erase voltages and then measuring the resulting threshold voltage shifts. As the number of cycles increases, the program and erase threshold levels begin to narrow, as shown in Figure 1-9. This degradation corresponds to poorer device performance, specifically longer program/erase times [19]. The reason for device degradation is high-voltage stress creating defects in the tunneling oxide, specifically oxide traps. These traps can assist tunneling conduction through the oxide and increase charge loss.

Obviously, charge loss for a non-volatile memory device is disastrous and minimizing charge loss is the ultimate concern. However, as devices scale down, the number of electrons that are stored on the floating gate significantly reduces, as shown in Figure 1-10. For example, at the 30nm design node, only about 100 electrons will charge the floating gate to correspond to a 6V threshold voltage shift. And considering less than a 10% tolerance for charge loss, that means over the course of 10 years only 10 electrons can be lost. Losing any more will make the memory device unusable. Memory cells are interconnected via wordlines and bitlines, as shown in Figure 1-4,
with standardized voltages applied to read, write, and erase the devices. If any one of the devices cannot retain charge to correspond to the standardized values, then that particular device in the array will be read incorrectly.

One of the main reasons for charge loss is defects in the tunneling oxide, either innately upon oxide growth or after significant voltage stresses, as mentioned earlier. The problem occurs when the oxide is made thinner, thus increasing the occurrence of defects. Defects in the oxide is damaging because they provide conductive pathways for charge to escape from the floating gate. This problem is exacerbated because the floating gate is a continuous, conductive layer where charges can freely travel to these charge sinks and thus cause significant charge loss. Although engineers would like to continue to scale down memory devices, in particular the oxide, to increase
density and performance (since tunneling probability increases with thinner oxide - faster write/erase times and lowerer program/erase voltages), they are limited by this retention problem. As a result, the tunneling oxide thickness is currently around 8nm.

If there were a solution to the charge retention and endurance problems, which one can pinpoint the tunneling oxide as the main culprit, then non-volatile floating gate memory devices can continue their miniaturization and increase performance trends.

1.3 Future of Flash Memory

The road ahead for flash memory will be difficult. The limitations regarding Flash, as discussed above, need to be resolved in order to continue Flash’s progress. Some have made predictions that floating gate technology will have to be replaced between the 25nm to the 45nm node [20]. However, a proposed roadmap for NAND Flash memories is shown in Figure 1-11.

![Roadmap for NAND Flash Memories](image)

Figure 1-11: Roadmap for NAND Flash Memories [14]

One possible solution to improve non-volatile floating gate memory scaling is to replace the continuous floating gate with discrete charge storage sites that make up the floating gate. Therefore, if any defects exist in the tunneling oxide below, only the charge storage sites in close proximity to that defect will lose its charge, thus localizing data loss. There are several ways that this can be done, two of which will be discussed in this thesis: Quantum Dot Non-Volatile Floating Gate Memory and Molecular Non-Volatile Floating Gate Memory.
Chapter 2

Metal-Oxide-Semiconductor (MOS) Capacitors

2.1 Introduction to MOS Capacitors

As stated in Chapter 1, a non-volatile floating gate memory device is a MOS transistor with a floating gate sandwiched in the dielectric between the control gate and the silicon substrate. However, as a first step in understanding possible floating gate substitutes (e.g. quantum dots or molecules), making a full-fledged MOS transistor is unnecessary. It is still possible to observe and understand the charging and discharging of the floating gate with a MOS capacitor. This is the approach taken in this paper, so a more thorough explanation of MOS capacitors is warranted.

2.1.1 MOS Capacitor in Thermal Equilibrium

A MOS capacitor is a capacitor made up of a metal top contact, an oxide as the dielectric, and a doped semiconductor substrate, onto which a back contact is made, as shown in Figure 2-1. The semiconductor in our experiments is p-type Silicon doped with Boron atoms at a concentration of $10^{15}$ cm$^{-3}$. Therefore, the majority carriers are holes with a concentration of $10^{15}$ cm$^{-3}$, equal to the doping concentration. Also, the minority are electrons with a concentration of
\[
\frac{n_o}{p_o} = \frac{n_i^2}{p_i} \\
\therefore n_o = \frac{10^{20}}{10^{15}} \\
\therefore n_o = 10^5 \text{ cm}^{-3}
\]

where \( n_i \) is the intrinsic electron and hole concentration at \( T=300K \) (\( 10^{10} \text{ cm}^{-3} \)), \( p_o \) is the hole concentration, and \( n_o \) is the electron concentration. The mass-action law is used to derive this equation [13, p.26].

A MOS capacitor is very similar to a traditional capacitor with two metal contacts sandwiching the oxide. However, by replacing one of the metal contacts with a semiconductor, the capacitance will vary with voltage and frequency, as discussed below.

At the flat-band voltage \( (V_{FB}) \), the MOS capacitor has no surface charge on either metal or semiconductor. Also, the work functions of the metal and semiconductor are aligned, as shown in Figure 2-2. For this discussion, we have set the flat-band voltage to be 0V. We depart from the equilibrium state when applying a bias to the device.
2.1.2 MOS Capacitor under Applied Bias

The MOS capacitor is made up of a series combination of the dielectric capacitance ($C_{ox}$) and the depletion capacitance ($C_s$). The dielectric capacitance is constant, $C_{ox} = \frac{\varepsilon_{ox} A}{t_{ox}}$ ($\varepsilon_{ox}$ is the oxide permittivity, $A$ is the device area, and $t_{ox}$ is the oxide thickness) while the depletion capacitance will vary with the applied voltage.

$$C_{total} = \frac{C_{ox} C_s}{C_{ox} + C_s}$$

Voltage is applied with the positive terminal at the top contact and the negative terminal at the bottom contact, as shown in Figure 2-1. The total capacitor acts like a voltage divider, so the applied voltage is divided across the oxide ($V_{ox}$) and the semiconductor ($V_s$)

$$V = V_{ox} + V_s$$

The voltage across the oxide is a function of total charge in the semiconductor ($Q_s$), $V_{ox} = \frac{Q_{ox}}{C_{ox}}$. As the semiconductor transistions through MOS regimes accumulation, depletion, and inversion, as shown in Figure 2-3, the amount of charge in the semiconductor ($Q_s$) will vary, and so will $V_{ox}$.

As a function of the voltage across the semiconductor ($V_s$), the semiconductor
will vary through accumulation, depletion, and inversion modes. As shown in Figure 2-4(a), accumulation is characterized by a build-up of majority carriers, holes for a p-type Si, at the surface. Assuming the flat-band voltage is equal to 0 ($V_{FB} = 0$), this occurs when a negative $V_s$ is applied. Using the energy band as in Figure 2-5(b), we see that with applied voltage the energy bands will bend upward at the surface. However, since we take the oxide to be a perfect insulator, there will be no current in the device. This translates to the Fermi energy levels staying flat, $\frac{dE_F}{dx} = 0$. But having the valence energy band ($E_V$) bend upward toward the Fermi level increases the hole charge density at the surface. Another way to explain it is with the equation

$$p(0) = p_o e^{-\frac{V_s}{kT}}$$

where $p(0)$ is the hole concentration at the surface and $p_o$ is the intrinsic hole concentration, as discussed in Section 2.1.1.

Depletion is characterized by majority carriers moving away from the surface, thus leaving a region depleted of mobile majority charges. This occurs when the applied voltage is positive, but not too positive (not exceeding $V_T$). Using the notation in the energy band diagram in Figure 2-2, depletion is when $\psi_{Bp} > V_s > 0$, which is when $p(0) > n(0)$. This time, the energy bands bend downward with the valence band moving farther away from the Fermi energy level. This reduces the concentration of holes at the surface of the semiconductor. The depletion region width varies with voltage as...
Figure 2-4: MOS Operation Modes [13]

\[ W_D = \sqrt{\frac{\varepsilon_{\text{semi}}^2}{C_{ox}^2} + \frac{2\varepsilon_s V}{qN_A} - \frac{\varepsilon_s}{C_{ox}}} \]

where \( V \) is the applied voltage and \( \varepsilon_s \) is the semiconductor permittivity.

Figure 2-5: MOS Energy Band Diagram [22]

As \( V_s \) is made more positive, the depletion region continues to grow until it reaches its maximum width of

\[ W_{D_{\text{max}}} = \sqrt{\frac{4eKT\ln(N_A/n_i)}{q^2N_A}} \]
As the energy bands continue to bend downward, the intrinsic level $E_i$ begins to cross the Fermi level. At that point, the hole and electron concentrations at the surface of the p-type semiconductor are equal. Past that point, the electron concentration exceeds the hole concentration and inversion occurs, as shown in Figure 2-4(c). As shown in Figure 2-5(c), the conduction band moves closer to the Fermi level and the electron concentration increases to exceed that of the intrinsic hole concentration at the surface.

Threshold voltage is defined as the voltage at which inversion occurs. This important variable is defined below

$$V_T = \frac{||Q_s||}{C_i} + 2\psi_{Bp}$$

where $||Q_s|| = \sqrt{2\varepsilon_s q N_A \psi_{Bp}}$.

In the accumulation region, the semiconductor capacitance ($C_s$) is high due to the accumulation of holes at the semiconductor surface. Therefore, the series combination gives a total capacitance approximately equal to just the oxide capacitance ($C_{ox}$).

$$C_s >> C_{ox}$$

$$C = \frac{C_{ox} C_s}{C_{ox} + C_s} \approx C_i$$

At the flatband voltage, in our case $V_s = 0$,

$$C_s = \frac{\varepsilon_s}{L_D}$$

where $L_D$ is the Debye length for holes and equals

$$L_D = \sqrt{\frac{\varepsilon_{semi} kT}{q^2 p_0}}$$
Remember, throughout the rest of the regimes, the oxide capacitance is constant. So calculating the total capacitance depends on adjusting $C_s$.

In the depletion region, the semiconductor capacitance is equal to

$$C_s = \frac{\varepsilon_{\text{semi}}}{W_D}$$

where $W_D$ is given above in Equation 2.1.2. As $W_D$ reaches its maximum value, equal to Equation 2.1.2, the semiconductor and total capacitances decreases to their minimum values, as shown in Figure 2-6.

As for the inversion regime, the capacitance depends on the frequency of the AC voltage signal. The charges in the semiconductor recombine and generate based on the speed of the signal. If the frequency is too high, the incremental charge creates a deep-depletion region which decreases the semiconductor capacitance even further. However, if the frequency is reasonable, inversion occurs as the incremental electron charge appears at the semiconductor surface. In this case, the semiconductor capacitance increases dramatically and thus the total capacitance approximately equals the oxide capacitance, as in the accumulation regime.
2.2 Experimental Setup

The MOS capacitors fabricated were measured using an Agilent 4294A Precision Impedance Analyzer and an Agilent 4156C Parameter Analyzer. The Agilent 4294A was used to measure capacitance as a function of voltage under a variety of frequencies. The equipment had a DC voltage range of ±40V and a frequency limit of 110Mhz [23]. Also, sweeps could have different speeds, from 1 (fastest, least accurate) to 5 (slowest, most accurate). Remember from Figure 2-6 that the sweep speed could affect CV measurements.

The Agilent 4156C was used to measure the leakage current in the capacitor. The measurement was made for a quasi-static CV measurement. A perfect capacitor would have no leakage current, but a realistic device does. If leakage currents were too high, this would indicate that a MOS floating gate capacitor would fail to program properly through tunneling, and would have poor retention.

The MOS capacitors are placed on a conductive gold chuck, which is set to ground. The backside of the MOS capacitors were coated with silver paste, after sanding away some oxide, to make an electrical contact with the gold chuck. For the top contact, a probe station was used. The actual probe was a Micromanipulator Compliance Probe. The compliance probe gave the best results in low frequency tests and proved the least destructive to the top contact of gold. Other probes like a gold wire were unstable during quasi-static CV tests and a regular probe would pierce through thin top metal contacts, thus destroying the device structure.
Chapter 3

Quantum Dot Floating Gate
Non-Volatile Memory

3.1 Introduction to Quantum Dot Floating Gate
Non-Volatile Memory

In conventional floating gate memory devices, a conductive, continuous polysilicon material is used to form the floating gate. However, limitations as discussed in Section 1.2 are preventing industry aims to miniaturize these memory devices even further.

A Quantum Dot (QD) memory, as shown in Figure 3-1, replaces the continuous polysilicon material with quantum dots, which overcomes the problem of inadvertent discharge through charge-leakage pathways in the insulating layer. An example of a similar device, but with Silicon nanoparticles, has been demonstrated [16]. Each quantum dot is a discrete charge storage site insulated from one another by an insulating shell. The shell creates a potential barrier preventing lateral conduction between quantum dots. Even if defects exist in the oxide, the poor lateral conductivity of QDs in the floating gate layer isolates that charge leakage in the vicinity of the defect [4]. The remainder of the charged QDs are unaffected and retains its memory.

The impact of using QD memories is that the insulator layer separating the semi-conducting QD core from the source-drain channel can be made thinner. As a result,
QD floating gate memory devices can be smaller in size, and operate faster and at lower voltages compared to present memory technology.

![Figure 3-1: Quantum Dot Memory](image)

3.2 Introduction to Quantum Dots

Quantum dots are zero-dimensional structures made of semiconductor materials with sizes ranging from 1-12nm. Since these nanoparticles are confined in 3-dimensions, there are some unique quantum effects. The energy levels within the quantum dots are discrete and tunable by adjusting the size of the quantum dot. This property can be interesting for electron and hole storage, as would be required in a Quantum Dot floating gate memory device.

Quantum dots are usually passivated with an inorganic shell with a wider bandgap. In addition to electrical insulation, the shell gives the quantum dot more robustness to withstand demanding conditions when integrated into other technologies [7].

Quantum dots are colloidal grown from different precursors, like Cd, Se, Zn, etc. Placed in a high temperature organic solvent, the quantum dots are formed from the nucleation of precursors. The reactions stop by lowering the temperature to room temperature. Finally, the quantum dots are deposited by spin-casting via phase segregation, which is discussed in Section 3.3.3.
3.3 Fabrication of QD Floating Gate Capacitor

The quantum dots used in fabricating the QD floating gate capacitor were 7nm ZnCdS and were obtained from Professor Moungi Bawendi’s group. The specific spectrum properties, usually interesting to scientists using quantum dots, were unimportant for this device. The key was to choose a core/shell quantum dot that would allow for charge storage and prevent lateral conduction between quantum dots.

The fabrication for the device, as shown in Figure 3-2, is discussed below.

![Device Structure of Quantum Dot Floating Gate Capacitor](image)

Figure 3-2: Device Structure of Quantum Dot Floating Gate Capacitor

3.3.1 RCA Clean

Boron doped, p-type, 10^{15} \text{ cm}^{-3} \text{ Silicon wafers were used as the substrate for the QD capacitors. The wafers were cleaned by undergoing an RCA clean in the Microsystems Technology Laboratory's (MTL) Integrated Circuits Laboratory (ICL). The RCA clean consists of several steps:}

1. Dip in H_2O_2:NH_4OH:DI H_2O for 10 minutes
2. Rinse in DI water
3. Dip in HF bath for 1 minute
4. Rinse in DI H_2O
5. Dip in H_2O_2:HCl:DI H_2O for 15 minutes
6. Rinse in DI H_2O

After being spun dry, the wafers were oxidized.
3.3.2 Oxidation

Thermal oxidation is a procedure using a furnace heated to a particular temperature with an $O_2$ ambient to grow a layer of dense $SiO_2$ layer. $O_2$ has a diffusivity through $SiO_2$ greater than Si’s through $SiO_2$. Therefore, $O_2$ diffuses through the oxide and reacts with Si. $SiO_2$ grows at the Si interface, rather than at the surface of the existing oxide, creating a clean $SiO_2$ layer [15]. The ability to grow clean, dense $SiO_2$ is one of the most important accomplishments in Silicon device fabrication.

$SiO_2$ growth follows the Deal-Grove Model, which says

$$x_{ox}^2 + Ax_{ox} = B(t + \tau)$$

where $x_{ox}$ is the thickness of the oxide, $t$ is the duration of oxidation (in hours), and $A$, $B$, and $\tau$ are constants dependant on oxidation temperature.

<table>
<thead>
<tr>
<th>Temperature ($^\circ$C)</th>
<th>$A$ ($\mu$ m)</th>
<th>$B$ ($\mu$ m$^2$/hr)</th>
<th>$\tau$ (hr)</th>
</tr>
</thead>
<tbody>
<tr>
<td>800</td>
<td>0.37</td>
<td>0.0011</td>
<td>9</td>
</tr>
<tr>
<td>920</td>
<td>0.235</td>
<td>0.0049</td>
<td>1.4</td>
</tr>
<tr>
<td>1000</td>
<td>0.165</td>
<td>0.0117</td>
<td>0.37</td>
</tr>
<tr>
<td>1100</td>
<td>0.090</td>
<td>0.027</td>
<td>0.076</td>
</tr>
<tr>
<td>1200</td>
<td>0.040</td>
<td>0.045</td>
<td>0.027</td>
</tr>
</tbody>
</table>

Table 3.1: Oxidation Constants [15]

A thermal oxide of 4nm is grown by oxidizing at 800$^\circ$C for 20min. However, using the Deal-Grove model in the linear regime (for small $x_{ox}$, $x_{ox} = \frac{B}{A}(t + \tau)$) and the constants above, we calculate an oxide thickness of 2.8nm under similar conditions. The reason is that the Deal-Grove model fails for thin oxides. It seems that the oxide growth is faster for thin oxides than the model calculates. Nonetheless, one can use the result above to calculate relative oxidation times for different thicknesses using

$$\frac{x_{ox1}}{t_1} = \frac{x_{ox2}}{t_2}$$

for small $x_{ox}$, where $x_{ox1}$ and $t_1$ are 4nm and 20min, respectively.
3.3.3 Spin-Casting Quantum Dots

The quantum dot floating gate layer is deposited via spin-casting, as shown in Figure 3-3. The actual quantum dots used are Blue 18, which are 7nm CdZnS dots.

Before spin-casting the actual quantum dot floating gate, quantum dot calibration needs to be done. Several solutions of quantum dots are made and spun to make several samples. Each sample is then analyzed under an Atomic Force Microscope (AFM) to determine the best solution with the best quantum dot coverage. By adjusting the quantum dot:solvent ratio, we can try to achieve a closed-packed monolayer of quantum dots. Calibrating the quantum dots to a 1:7 ratio produced a monolayer as shown in Figure 3-4.

Figure 3-3: Quantum Dot Deposition via Phase Segregation [5]

Figure 3-4: Atomic Force Microscope Image of Quantum Dot Calibration (1:7 QD:Chloroform)
3.3.4 Parylene Deposition

Parylene, a polyxylene polymer with high dielectric strength, was used as the gate dielectric. Depositing a thick film of parylene can produce a uniform, pinhole free, conformal layer. It has a dielectric constant comparable to SiO$_2$, (3.9 for SiO$_2$ and 3.1 for Parylene-C). There are several variants of parylene available: Parylene-N, Parylene-C, and Parylene-D, as shown in Figure 3-5 [18]. Parylene-C was used for this device based on extensive use by research colleagues.

![Chemical Structure of Parylene Variants](image)

Figure 3-5: Chemical Structure of Parylene Variants

Parylene-C is deposited using chemical vapor deposition (CVD) in a dedicated Parylene chamber in MTL's Technology Research Laboratory (TRL). To determine how much material is needed, use the ratio for Parylene C:

$$0.2g = 2000 \text{ Å}$$

Therefore, for the 1350 Å of Parylene that was grown, about 0.15g of material was used. The parylene growths may not be the same from growth to growth even though the amount of material may be the same. Therefore, nanometer precision in thickness control is unachievable using this method of Parylene deposition.
3.3.5 Gold Deposition

Gold deposition was done using the metals evaporator in Professor Vladimir Bulovic's Laboratory of Organic Optics and Electronics. Only 400 Å of gold was deposited, which is too thin to prevent complete penetration by the testing probes. Therefore, a gold wire was silver pasted to the probe. This gentle approach proved to be adequate. However, in future experiments, depositing Chromium (50 Å) and then Gold (600 Å) proved to be best. A shadow mask, as shown in Figure 3-6, was used.

![Figure 3-6: Shadow Mask for Top Metal Contact (Diameters: A=1mm, B=0.5mm, C=0.2mm)](image)

3.4 Experimental Results of QD Capacitor

Measuring the Quantum Dot Floating Gate Memory device with the Agilent 4294A for capacitance with respect to voltage results in Figure 3-7. As we see in the plot, there is hysteresis with a voltage window of approximately 1.5V. Taken that both holes and electrons are stored in the floating, as explained below, let us assume that the voltage shift for each carrier is 0.75V.

Using the equation, \( V_T = V_{T0} - \frac{\Delta Q}{C_{ox}} \), we can calculate the number of charges stored on the floating gate. To solve for \( Q \), we use \( \Delta V_T = -\frac{\Delta Q}{C_{ox}} \) and the value for \( C_{ox} \). We can calculate \( C_{ox} \) by using

\[
C_{ox} = \frac{C_{i1}C_{i2}}{C_{i1} + C_{i2}}
\]

39
where $C_{11}$ is the capacitance of the gate dielectric, and $C_{12}$ is the capacitance of the tunneling dielectric. Using the standard equation for capacitance,

$$C = \frac{\varepsilon A}{d}$$

we can calculate the capacitance for each dielectric. The dielectric constant for parylene and SiO$_2$ are 3.1 and 3.9, respectively. Also, taken that the top metal contact has a diameter of 1mm, as shown in Figure 3-6, then the area is $A = \pi \left(\frac{1\text{mm}}{2}\right)^2$. Therefore,

$$C_{11} = 162\text{pF}$$
$$C_{12} = 6.8\text{nF}$$

$$\therefore C_{ox} = 162\text{pF}$$

As shown in Figure 3-7, we see that $C_{ox} = 410\text{pF}$. Taking the experimental result of $C_{ox}$, instead.

$$Q = \Delta V C_{ox} = (0.75V)(410\text{pF})$$
$$Q = 3 \times 10^{-12} = nq = n(1.6 \times 10^{-19}C)$$

$$\therefore n = 2 \times 10^9$$

Therefore, we have 2 billion electrons stored in the quantum dot floating gate layer. Given that each quantum dot is 7nm in diameter and the floating gate has an effective diameter of 1mm, we calculate that there are 15 billion storage sites. We are assuming that the quantum dots are completely packed on the floating gate, however, this may not be true. Nonetheless, we see that not all the charge storage sites are charged.
The hysteresis is indicative of charge storage, and by looking at the direction of the $V_T$ we can determine the type of charge stored. As seen in Figure 3-7, there is a right shift in $V_T$ after applying -5V and a left shift in $V_T$ after applying +5V. Importantly, because there isn’t a uncharged state except for the initial sweep (not shown), then there is both hole and electron charging/discharging.

![Figure 3-7: CV Measurement of QD Device](image)

A right $V_T$ shift indicates that electrons are charged onto the floating gate. One way to understand that is by referring to

$$V_T = V_{T0} - \frac{Q}{C_{ox}}$$

. Because $V_T$ shifts to the right, or is more positive, Q has to be negative - electrons.

Another way to understand this is by thinking about the effect the stored charge has on the electric field in the device and the semiconductor regimes. For example, at -5V this p-type semiconductor substrate is in accumulation. When the voltage is swept in the positive direction, the electric field from the top metal contact begins to repel the holes in accumulation and creates a depletion region. However, we see that
the curve shifts to the right, thus indicating that more positive voltage is required to induce depletion. This is indicative of electrons being stored in the floating gate, as the electrons shield and minimize the electric field that is felt by the semiconductor. Thus, a more positive voltage is required.

 Appropriately, a left $V_T$ shift indicates that holes are charged onto the floating gate. Using the same analysis above, this phenomenon can be understood. First, $V_T$ shifts to the left, or is less positive, so $Q$ has to be positive - holes. Second, because the holes on the floating gate shield the relatively less positive electric field and repel any holes from accumulating at the semiconductor, a more “negative” voltage is required.

In summary, at -5V when the p-type semiconductor is in accumulation, we have electron charging of the floating gate. Also, at +5V when the p-type semiconductor is in depletion/inversion, we have holes charging the floating gate. However, this does not make sense if the charging is a result of carriers tunneling onto the floating gate, as the charges that exist on the semiconductor surface are the opposite type that what is actually being charged. We suspect that the charges are coming from the top metal contact and injecting through the parylene insulator.

Charge retention tests were also done by charging and then reading the device trying to avoid recharging the device. To determine electron retention, we charged the device at -5V and swept the device from 0V to -2V at different times. As shown in Figure 3-8(a), we see the CV curves drift back left from the initial right charge shift as electrons are lost. Also, as shown in Figure 3-8(b), we see the CV curves shift back right after the initial left charge shift as holes are lost.

Therefore, we see in Figure 3-9 that the electron and hole retention times are 1200s and 2000s, respectively. These retention times fall short of the 10 year retention time requirement for Flash memory. The reason for this result, we believe, is that the gate dielectric is leaky, so much so that the charges are injected through it to charge the floating gate. As carriers are able to charge through defects in the gate dielectric, they can discharge through the defects as well. Obviously, this is a problem that needs to be rectified if a fully functional quantum dot floating gate memory is to be realized.
A method for fabricating ideal quantum dot floating gate memories remains elusive as the tools for generating the needed dense nanopatterns on the scale of 2nm to 10nm, have not yet been demonstrated. Currently, QDs are typically deposited via spin-casting and as a result, there is variation in QD density - specifically order, translation, and rotation.

Due to current QD deposition techniques (i.e. spin casting, self-assembly) the
number of charge storage sites can vary quite a bit between memory cells. The QD layer can vary in order, translation, and rotation, all of which alters the number of charge storage sites between the gate and the channel, which alters the shift in threshold voltage, which ultimately alters the accurate operation of the memory device.

Two assumptions we made are that all the QDs are of uniform size and the tunneling distances are equidistant. These assumptions allow us to ignore the variation in tunneling probabilities that would result if these assumptions were not made, thus allowing us to focus on QD layer variation in order, translation, and rotation.

Joushua Leu (student with Prof. Karl Berggren) and I simulate (in Matlab) the QD layer’s variation in order, translation and rotation to determine requirements on the size of QDs to satisfy acceptable consistency across memory cells.

### 3.5.1 Simulations

We approach this simulation by considering two possible QD deposition patterns: completely random and perfect hexagonal packing. For each of these patterns, a square covering the QDs represents the control gate of the memory cell, as shown in Figure 3-10. When counting the number of charge storage sites, we count the number of QDs that exist below the square.
Figure 3-10: (left) Unordered Monolayer of QDs. (right) Hexagonally Closed-Packed Ordered monolayer of QDs. The QDs are part of the device if encompassed by the top gate contact, which is represented by the square.

We vary the relative rotational and transitional positions between the plane of quantum dots and gate. By varying the gate within a parallelogram about a single QD and rotating over 360 degrees, we sampled over 10,000 variations of quantum dot and gate placements. This allows us to simulate the uncontrollability in quantum dot processing.

At each instance of the relative rotational and translational position, we calculate the number of quantum dots covered by the gate. We have defined a covered quantum dot by its center located within the edges of the square. This definition requires that at least half of the quantum dot is covered, which is a satisfactory definition for all quantum dots except those in the corners of the square. Note that this exception is ignored.

To assess these results, we define three quantities.

Variation,

\[ V = \frac{Maximum \, \# \, of \, QDs - Minimum \, \# \, of \, QDs}{Average \, \# \, of \, QDs} \]

Square size length QD size,

\[ R = \frac{Square \, Length}{Quantum \, Dot \, Diameter} \]

\[ E = Number \, of \, Quantum \, Dots \, Encompassed \, with \, R = 11.6 \]
Our simulation was conducted on ordered and unordered quantum dots. The random quantum dot array was obtained from actually spin-casting quantum dots, taking its SEM, and using our own custom software to detect their locations, Figure 3-11. Using the locations, we ran our simulation and obtained the variation curve (V vs. R), Figure 3-12. We see that to limit the variation to +/-10% and +/-5%, the necessary R are 13 and 21, respectively. These critical values of R reveal the required quantum dot sizes once the gate dimension is chosen for a particular design node.

![Image of Quantum Dot array](image)

Figure 3-11: Scanning Electron Microscope (SEM) image of Quantum Dot array. Red region indicates QD locations detected by software. Simulated analysis is based on detected locations from SEM images.

The size requirements can be made less demanding if quantum dots were ordered, like in a hexagonal close packed plane. Once again, an actual ordered quantum dot monolayer sample was made, Figure 3-13, from which the locations were detected by software and used for our simulation. The results of the simulation are shown in Figure 3-14. We see that to limit the variation to +/-10% and +/-5%, the necessary R values are 8 and 14, respectively. Thus, having an ordered plane provides a significant improvement over an unordered plane.

With a perfectly ordered monolayer of quantum dots, the next step in achieving less demanding quantum dot size requirements is to allow for 1-D grain alignment.
Figure 3-12: Random Monolayer of QDs with Translation and Rotation. Square side length-QD size ratio (R) of 8 and 14 are needed to obtain variations (V) of +/-10% and +/-5%, respectively.

Figure 3-13: Scanning Electron Microscope (SEM) image of ordered Quantum Dot array. QD locations detected by software. Simulated analysis is based on detected locations from SEM images.
Figure 3-14: Ordered Monolayer of QDs with Translation and Rotation variability. Square side length-QD size ratio (R) of 8 and 14 are needed to obtain variations (V) of +/-10% and +/-5%, respectively.

For example, we could align one side of the gate with a grain in the QD monolayer, yet still allow for translational variation in the perpendicular direction. However, as seen in Figure 3-15, this approach results in significant discontinuous coverage variation as a whole row of QDs is removed from the coverage area under the gate. Nonetheless, we see that size requirements are a little less stringent - variation of +/-10% and +/-5% requires R of 7 and 13, respectively.

### 3.5.2 Conclusion

As industry continues to pursue smaller memory devices, it is necessary to determine the requirements on QD storage layers. Ultimately, these are requirements on the QD size at particular design nodes, summarized in Table 3.2. The calculated QD sizes will be difficult to achieve because today’s quantum dots are at best around 1nm in diameter [4] and still need another 1nm for the insulating cap.

Our simulations of ordered and unordered quantum dot monolayers show that QD memory devices will require significant advancement in QD patterning and QD fabrication in order to make QD memory a viable long-term replacement to Flash
memory. If quantum dots cannot be deposited with increased order and translation-rotation control, then the QD size requirements will continue to be challenging.

One possible solution is to replace the Quantum Dot floating gate with a molecular film - PTCDA, for example. In this structure, 1nm molecules of PTCDA become perfect nanostructured sites of charge storage. The polycrystalline arrangement and relatively smaller intermolecular spacing creates for consistent molecular packing. Add to that poor lateral conduction, and the molecular floating gate memory may provide a reliable alternative to QD-FGMs.

<table>
<thead>
<tr>
<th>Technology Design Node</th>
<th>V = +/- 10%</th>
<th>V = +/- 5%</th>
</tr>
</thead>
<tbody>
<tr>
<td>50nm</td>
<td>7nm</td>
<td>4nm</td>
</tr>
<tr>
<td>39nm</td>
<td>6nm</td>
<td>3nm</td>
</tr>
<tr>
<td>28nm</td>
<td>4nm</td>
<td>2nm</td>
</tr>
<tr>
<td>20nm</td>
<td>3nm</td>
<td>1nm</td>
</tr>
</tbody>
</table>
Chapter 4

Molecular Floating Gate Memory

4.1 Introduction to Molecular Floating Gate Non-volatile Memory

One way to overcome the QD-to-QD size challenge, as discussed in Section 3.5.2, is to replace QDs with molecules as the floating gate. Ideally, these molecules would be approximately 1nm² in size and have poor lateral conduction to prevent complete charge loss due to defects in the tunneling oxide. In essence, the molecules would be extremely small charge storage sites. Conveniently, molecular films have the advantage of inherent order and intermolecular spacing [9] which could overcome the technological challenges of device miniaturization associated with QDs. The structure for a molecular floating gate non-volatile memory is shown in Figure 4-1.

Figure 4-1: Molecular Floating Gate Memory Transistor
4.2 Introduction to Organic Molecular Materials

In choosing a molecular material as the floating gate, we want an easily evaporated material with favorable physical and electrical properties. For instance, a planar polycrystalline material with poor lateral mobility is desirable. Poor lateral mobility is important, as it was with a QD floating gate, so that if there is a defect in the tunneling oxide, the charge leakage is localized around that defect.

We use 3,4,9,10-perylenetetracarboxylic dianhydride (PTCDA) as the molecular film in fabricating the floating gate memory capacitor. PTCDA is a planar molecule that forms a herring-bone molecular packing structure with planes of the molecules oriented parallel to the substrate [11], as shown in Figure 4-3. In an ordered monolayer of PTCDA molecules, a single molecule occupies an area of approximately 1 nm² [12], as shown in Figure 4-2. Compared to QDs, which typically exhibit size and order variability, molecular films have the highly desirable size and morphological consistency, which provides relative similarity in the electronic energy levels. Using a Scanning Tunneling Microscope (STM), which is a technique to achieve atomic level resolution of a material, an image of PTCDA is shown in Figure 4-4 [8].

![PTCDA Structure](image)

Figure 4-2: Molecular Structure of 3,4,9,10-perylenetetracarboxylic dianhydride (PTCDA) [10]

The charge transport properties of PTCDA are important to floating gate memories. These properties depend on among other things, molecular ordering, orbital overlap, and molecular orbital energy levels [17]. Carriers have relatively high mobility perpendicular to the molecular planes, which have significant pi-bond overlap [11]. However, carrier mobility parallel to the molecular plane is 4 orders of magni-
Figure 4-3: Crystal Structure of 3,4,9,10-perylene-tetracarboxylic dianhydride (PTCDA) [10]

Figure 4-4: Scanning Tunneling Microscope of 7 x 7 nm² Monolayer of PTCDA on Ag [8]
tude smaller - $\mu = 0.1 \text{ cm}^2/\text{Vs}$ in the stacking direction and $\mu = 10^{-5} \text{ cm}^2/\text{Vs}$ in the planar direction [12]. The low lateral mobility of electrons provides the same advantage as QDs in floating gate memories: if a defect exists in the tunneling oxide below the floating gate, charges are unlikely to transport laterally to discharge through that defect.

Below, we demonstrate the use of PTCDA molecular organic thin films as nanos-structured, programmable charge storage films in capacitive structures.

## 4.3 Fabrication of Molecular Floating Gate Capacitor

In order to demonstrate the programmability of PTCDA floating gate memory devices, we start by constructing and understanding the PTCDA floating gate capacitor, as shown in Figure 4-5. We begin our fabrication process with a p-type Si substrate - Boron at $10^{15} \text{ cm}^{-3}$.

![Figure 4-5: Device Structure of a PTCDA Capacitor](image)

The fabrication steps are:

1. RCA clean Silicon wafers (as discussed in Section 3.3.1)
2. Dry oxidation at 800C to grow 4 nm SiO2 (as discussed in Section 3.3.2)
3. Evaporation of 50nm PTCDA
4. Plasma Enhanced Chemical Vapor Deposition (PECVD) of SiO2 - 150nm thick
5. Gold Deposition as top-gate contact (as discussed in Section 3.3.5)
6. Silver paste bottom-gate contact
4.3.1 Evaporation of PTCDA

PTCDA is deposited on the SiO$_2$ substrate via organic evaporation. The organic evaporator is found in Professor Bulovic's Laboratory of Organic Optics and Electronics (LOOE) and is dedicated to 6 organic materials, with PTCDA being one of them. The organic evaporator works by applying power to a crucible containing PTCDA. The power heats the organic material to the sublimation temperature, at which point the PTCDA material will evaporate and deposit onto the substrate in the chamber. For PTCDA, that sublimation temperature is found in literature to be approximately 450°C [11]. However, experimentally, we found that an unencapsulated film of PTCDA on a device would evaporate at temperatures as low as 250°C. This is something that needs to be further explored as the project moves forward.

Using the parameters indicated in the organic evaporator logbook, the rate of PTCDA deposition was 2Å/s. It is shown in [10] that there is a relationship between the deposition rate and molecular order for PTCDA. As we increase the deposition rate, the molecular order will also increase. As a result, the carrier mobility in the stacking direction will increase as shown in Figure 4-6. Fortunately, this increase in mobility is limited to the stacking direction and not in the planar, lateral direction.

![Deponation Rate vs Mobility](image)

Figure 4-6: Carrier Mobility in PTCDA as a Function of Deposition Rate [10]

After depositing 50nm of PTCDA, an AFM image was taken to examine the
structural features of the material, as shown in Figure 4-7. The PTCDA surface is not perfectly flat, which may make encapsulating it, a future step that will be explored in Section 5.1, difficult. We need an encapsulation material that will easily conform to the PTCDA surface.

4.3.2 Plasma Enhanced Chemical Vapor Deposition (PECVD) of SiO₂

In Chapter 3, we presented a QD floating gate capacitor with parylene as the gate dielectric. Though parylene was a very good pin-hole free insulator when thicker than 1500Å, it was very difficult to achieve good quality insulators at thinner thicknesses. Also, as we consider molecular floating gate memory as a replacement to conventional Flash memory, we need to minimize unnecessary fabrication process changes that industry would have to make. For these reasons, SiO₂ was used as the gate oxide.

There are several methods to deposit SiO₂. As discussed in Section 3.3.2, we can undergo thermal oxidation, but at temperatures at least 800°C. Such high tempera-
tures is destructive to the PTCDA layer that is exposed during the oxidation step. Remember, PTCDA evaporates at much lower temperatures (i.e. as low as 250°C). Therefore, we need a low-temperature alternative that can allow us to deposit a good quality oxide layer but without destroying the exposed PTCDA film.

Plasma Enhanced Chemical Vapor Deposition (PECVD) is great alternative for low-temperature oxidation. PECVD is able to achieve reactions as low as 250°C. Such low-temperature oxidation is done by having a plasma enhance the reaction of Silane (SiH₄) and Nitrous Oxide (N₂O). The two gases flow into the PECVD chamber where electrical power is used to energize the gas mixture. The Silane and Nitrous Oxide collide with such great energy that it transforms the gas molecules into reactive species. These species react at the surface of substrate to form SiO₂. Using PECVD, pinhole-free SiO₂ films as thin as 300Å can be achieved without destroying the PTCDA substrate.

Spin-on-Glass (SOG) is another alternative to growing oxides at low-temperature. A solution of SiO₂ is spun on the substrate and then baked to evaporate the solvent. Though this approach was not explored because of concerns of the destructive nature of the solvent on our organic substrate, it might be something to keep in consideration.

4.4 Experimental Results of Molecular Floating Gate Capacitor

We tested our PTCDA floating gate capacitors using an Agilent 4294 Impedence Analyzer at 10kHz test frequency and voltage sweeps of +/-10V applied to the Au electrode and the Silicon substrate grounded. The resulting C-V curve, as shown in Figure 4-8, shows hysteretic behavior with a 10V Vₜ shift, which is indicative of charging/discharging of the PTCDA layer. The dip in capacitance at positive voltage is due to the fast frequency (100kHz) and fast sweeps at which the device was measured, thus causing deep-depletion, as explained in Section 2.1.2.

Just as was the case with the QD floating gate capacitors, the charging mechanism
is not tunneling from the silicon substrate. We believe that charge carriers are injected into the floating gate from the top electrode.

![Figure 4-8: Capacitance vs. Voltage for a PTCDA Floating Gate Capacitor Device at 100kHz frequency](image)

We determined the charging mechanism by analyzing the voltage shifts in the CV curves, as we did for the QD floating gate capacitors. A thorough explanation is given in Section 3.4 with the energy band diagram of the structure shown in Figure 4-9. To recap, a left voltage shift is indicative of hole charging and a right voltage shift is indicative of electron charging. This is what we see in Figure 4-8, which is not what we want. We want the hole charging to occur through tunneling at the point when the p-type semiconductor is in accumulation and electron charging when the p-type semiconductor is in inversion. We believe the carriers are injected through the gate dielectric, which is a problem that needs to be resolved as we improve the device in the future.

We see that the hysteresis window of the device is approximately 10V. We can estimate that each carrier is causing a 5V shift away from the CV curve of the uncharged device (not shown), from which we can determine the number of charges that are stored on the PTCDA floating gate. As presented in Section 3.4,

\[
C_{ox} = \frac{C_{11}C_{12}}{C_{11} + C_{12}}
\]

where \(C_{11}\) and \(C_{12}\) are the capacitances of the gate dielectric and tunneling dielectric,
respectively. From the equation $C = \varepsilon \frac{A}{d}$, we get

$$C_{i1} = 47pF$$

$$C_{i2} = 1.7nF$$

$$.\therefore C_{ox} = 46pF$$

The device has an area $A = \pi \left(\frac{0.5mm}{2}\right)^2$ and dielectric constants for the PECVD SiO$_2$ and the thermal SiO$_2$ are 4.1 and 3.9, respectively [6]. We see that the derivation of $C_{ox}$ matches experimentally, as shown in Figure 4-8.

Lastly, we solve for $Q$ from the equation $V_T = V_{T0} - \frac{Q}{C_{ox}}$:

$$Q = \Delta V_TC_{ox} = (5.0 V)(46pF)$$

$$Q = 230 \times 10^{-12} = nq = n(1.6 \times 10^{-19} C')$$

$$.\therefore n = 1.4 \times 10^9$$

We calculate that there are approximately 1.4 billion carriers stored in the floating gate. Given that each molecule is approximately 1nm$^2$ in area, we calculate that the
number of charge storage sites in the device is

\[ n \times (1nm)^2 = \pi(0.5nm/2)^2 \]

\[ \therefore n = 196 \times 10^9 \]

We see that just a small percentage of PTCDA storage molecules are filled with carriers. It is not required that all the storage sites should be filled, but it is required that there is voltage shift consistency across devices, and there is, as shown in Figure 4-8.

Endurance tests were also done to understand the reliability of the device. This was done by programming the device at +10V and -10V over many cycles. As the device degrades, the hysteresis window will narrow because defects appear in the oxides, thus providing leakage pathways. However, as shown in Figure 4-10, the device withstands over 300,000 cycles.

Since the charging mechanism is not via tunneling through the tunneling dielectric from the semiconductor, we were not ready to proceed in conducting retention tests. The fact that carriers are able to travel through the gate dielectric is a problem that needs to be resolved before we can obtain reasonable retention tests.

We also looked at n-type silicon devices to show that this device works just the
same no matter the type of doping. However, note that the same MOS device regimes exist for an n-type semiconductor, but with the majority and minority carriers being electrons and holes, respectively.

It is important to note that the devices should show area proportionality because of the fundamental equation for capacitance. As shown in Figure 4-11(a), the capacitance in accumulation is equal to approximately 250pF. Given that the device in Figure 4-11(b) has half the diameter, and therefore a quarter of the area, the device should have a quarter of the capacitance in device Figure 4-11(a), which is approximately true. This gives us confidence that the dynamics of the device programmability are due to the floating gate and will be proportional to the area of the device.
4.5 Conclusion

We have shown that molecular floating gate memory devices have significant potential to replace conventional flash memory devices. We have shown the charging/programmability of the molecular film, PTCDA, its repeatability across many devices, and its long endurability. We have also shown that the type of semiconductor is irrelevant, as both types of carriers can be stored in the molecular film. We believe that the charging mechanism is by injection through the gate dielectric, but we are working towards fabricating a floating gate capacitor that programs by carriers tunneling from the substrate. Though much remains to be answered, like the extent of charge retention and stability, more experiments will be conducted to understand those issues. However, these results show a first step towards a possible approach to memory miniaturization by using molecular floating gate memories.
Chapter 5

Future Work

We should explore the use of a protective layer to the PTCDA layer to help us achieve a fully functional molecular floating gate memory capacitor. Once that is accomplished, we can then pursue a memory transistor, and ultimately a simple, operational memory device.

5.1 Protective Layer

A protective layer is a thin insulating film that would protect PTCDA from extreme subsequent processing conditions and allow for the use of fabrication equipment that would otherwise be off limits because of organic contamination.

We have seen in experiments that when devices with PTCDA are annealed at 250°C, PTCDA begins to evaporate off of the device. We anneal the devices because it improves the dielectric quality - fewer defects. A protective layer would encapsulate the PTCDA layer and prevent it from evaporating away, thus keeping the floating gate intact.

The protective layer would also allow us to use certain fabrication equipment that would produce high quality low temperature oxides. Without the protective layer, the organic material, PTCDA, would be exposed and could contaminate the equipment. Contaminated equipment is detrimental to other people's work and all precautions are taken to prevent that from occurring. A completely encapsulating layer could
resolve this problem.

So far, we have looked at using Al₂O₃ as a possible protective layer. There are two ways we can achieve this: 1. Evaporate a thin film of Al and have it oxidize in ambient and 2. Sputter a thin film of Al₂O₃. The first option is difficult to do because Al self-terminates oxidization at about 2-5nm. Therefore, we would have to evaporate an Al film less than 5nm thick. The second option is also difficult because sputtering a thin film of Al₂O₃ is difficult. As shown in Figure 5-1 and 5-2, 5nm and 10nm of Al₂O₃ do not have a completely smooth encapsulation of the PTCDA underneath.

![AFM of 5nm Al](image)

**Figure 5-1: AFM of 5nm Al**

Other protective layers should be explored to achieve an encapsulating, protective layer that would allow us to complete a molecular floating gate capacitor.
Figure 5-2: AFM of 10nm Al
Chapter 6

Conclusion

Flash memory is nearing the end of the miniaturization road because of reliability concerns rooted in the tunneling dielectric and the floating gate. We have shown that replacing the continuous, conductive floating gate of conventional Flash memory devices with quantum dots is a reasonable option. However, we have found through simulation that as we pursue smaller technology nodes, quantum dot to quantum dot spacing needs to be as small as 1nm to limit device variation. This limitation has led us to the use of molecules to make up the floating gate. With poor lateral conduction, we can achieve a controllable floating gate memory device. There is still more work to be done, for example a protective layer to protect the molecular floating gate. Nonetheless, we have shown the potential for molecular floating gate memory devices to replace conventional Flash memory and overcome the limitations to miniaturizing memory devices in the future.
Appendix A

Quantum Dot Size Simulation

Code

This is a partial set of the code written with Joshua Leu, who works with Professor Karl Berggren.

```matlab
function n = inside(E,F,A,B,C,D)
    %n=number of points inside square
    %E:x coord of points in lattice
    %F:y coord of points in lattice
    v1x = A(1,1) - E;
    v1y = A(1,2) - F;
    v2x = B(1,1) - E;
    v2y = B(1,2) - F;
    v3x = C(1,1) - E;
    v3y = C(1,2) - F;
    v4x = D(1,1) - E;
    v4y = D(1,2) - F;
    a = mag(v1x, v1y);
    b = mag(v2x, v2y);
    c = mag(v3x, v3y);
```
d=mag(v4x,v4y);

% dot products
e=v1x.*v2x+v1y.*v2y;
f=v2x.*v3x+v2y.*v3y;
g=v3x.*v4x+v3y.*v4y;
h=v4x.*v1x+v4y.*v1y;

% matrix of thetas
theta1=acos(e./(a.*b));
theta2=acos(f./(b.*c));
theta3=acos(g./(c.*d));
theta4=acos(h./(d.*a));

% single matrix of theta
theta=theta1+theta2+theta3+theta4;

% matrix of booleans
theta=theta-2*pi;
l=(theta.^2)<1e-6;

% number of points inside square
n=sum(sum(l));

function mag = mag(A,B)
mag=sqrt(A.^2+B.^2);

function [refpt]=ref(C,D)
ax=C(150,150);
ay=D(150,150);
bx=C(150,151);
by=D(150,151);
cx=C(151,151);
cy=D(151,151);
dx=C(151,150);
dy=D(151,150);
% \texttt{ax} = C(1,1);
% \texttt{ay} = D(1,1);
% \texttt{bx} = C(1,2);
% \texttt{by} = D(1,2);
% \texttt{cx} = C(2,2);
% \texttt{cy} = D(2,2);
% \texttt{dx} = C(2,1);
% \texttt{dy} = D(2,1);
\texttt{randx} = 3*\texttt{rand} + \texttt{ax};
\texttt{if randx} < (1 + \texttt{ax})
\texttt{i} = 0;
\texttt{while} \texttt{i} == 0
\texttt{randy} = 2*\texttt{rand} + \texttt{ay};
\texttt{if randy} <= (\texttt{ay} + \texttt{sqrt}(3)*(\texttt{randx} - \texttt{ax}))
\texttt{i} = 1; \texttt{refpt} = [\texttt{randx}, \texttt{randy}];
\texttt{else} \texttt{i} = 0;
\texttt{end}
\texttt{end}
\texttt{elseif randx} < \texttt{bx}
\texttt{i} = 0;
\texttt{while} \texttt{i} == 0
\texttt{randy} = 2*\texttt{rand} + \texttt{ay};
\texttt{if randy} <= \texttt{cy}
\texttt{i} = 1; \texttt{refpt} = [\texttt{randx}, \texttt{randy}];
\texttt{else} \texttt{i} = 0;
\texttt{end}
\texttt{end}
\texttt{else i} = 0;
\texttt{while} \texttt{i} == 0
\texttt{randy} = 2*\texttt{rand} + \texttt{ay};
\texttt{end}
if randy >= (cy - sqrt(3)*(cx - randx)) & randy <= cy; 
i = 1; refpt = [randx, randy]; 
else i = 0; 
end 
end 
end 

function [A, B, C, D] = mysquare(refpt, angle, length) 
% rotate counterclockwise 
RS = length / 2 * [cos(angle) - sin(angle); sin(angle) cos(angle)]; 
A = [1 1]'; 
B = [-1 1]'; 
C = [-1 -1]'; 
D = [1 -1]'; 
A = refpt + RS*A; 
B = refpt + RS*B; 
C = refpt + RS*C; 
D = refpt + RS*D; 
A = A'; 
B = B'; 
C = C'; 
D = D'; 

function [m] = singlesample(E, F, refpt, 1) 
%m: vector of n for each angle 
%set refpt, determine if each lattice point is inside square 
%each round has a new refpt 
m = []; 
for j = 1:90 
    angle = pi / 180 * j; 
    [A, B, C, D] = mysquare(refpt, angle, 1); 

x = inside (E, F, A, B, C, D):
m = cat (2, m, x):
end

function [per] = manysample (1)
n = 230;
[E, F] = hcplattice (n);
refpt = ref (E, F);
X = [];
for i = 1:10
    m = singlesample (E, F, refpt, 1);
    X = cat (2, X, m);
end
% hist (X)
sig = std (X);
m = mean (X);
p = numel (X); % total number of samples
c = 0; % count of elements within +/- 1 sig
for i = 1:p
    if X (1, i) <= (m + sig) && X (1, i) >= (m - sig)
        c = c + 1;
    end
end
per = (p - c) / p;

function mysimulation
    clear all
% for i = 1:100
%     y (1, i) = manysample (i);
% many samples for each length, y is vector of %'s
% end
y(1,1) = manysample(1):
y(1,2) = manysample(25):
y(1,3) = manysample(50):
y(1,4) = manysample(75):
y(1,5) = manysample(100):
x = [1, 25, 50, 75, 100];
plot(x, y)
Bibliography


