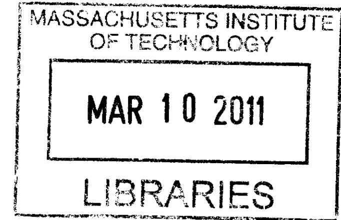
Organic Molecular Floating Gate Memories

by

Sarah Paydavosi

B.S. Electrical Engineering
University of Tehran, 2005

M.S. Electrical Engineering
University of Tehran, 2007


MASSACHUSETTS INSTITUTE OF TECHNOLOGY
MAR 10 2011
LIBRARIES

ARCHIVES

Submitted to the Department of Electrical Engineering and Computer Science in partial fulfillment of the requirements for the degree of

Masters of Science in Electrical Engineering and Computer Science

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

February 2011

© Massachusetts Institute of Technology 2011. All rights reserved.

Signature of Author.............................
Department of Electrical Engineering and Computer Science
January 28, 2011

Certified by ....
Vladimir Bulović
Professor of Electrical Engineering and Computer Science
Thesis Supervisor

Accepted by ....
Terry Orlando
Chairman Department Committee on Graduate Students

Organic Molecular Floating Gate Memories
by

Sarah Paydavosi

B.S. Electrical Engineering
University of Tehran August 2005

M.S. Electrical Engineering
University of Tehran August 2007

Submitted to the Department of Electrical Engineering and Computer Science
February 2011, in partial fulfillment of the
requirements for the degree of
Masters of Science in Electrical Engineering and Computer Science

Abstract

Flash memory devices dominate the non-volatile memory market, with device structures that utilize charge storage in polysilicon floating gates imbedded in insulating silicon oxide films[1]. As demands for high storage density, high chip memory capacity, and decreasing process costs continue to mount, conventional flash memory has found it challenging to continue scaling and it may reach fundamental scaling limits because of the minimum tunnel oxide thickness and poor charge retention due to defects in the tunneling oxide, necessitating modification in the implementation of the flash memory technology[2]. In this study nano-segmented floating gate memories consisting of a uniform set of identical organic dye molecules were fabricated and evaluated for potential use as programmable charge storage and charge retention elements in a future flash memory technology. Viability of molecular thin films to serve as an energetically-uniform set of ~1nm in size charge- retaining sites is tested on a series of molecular materials, the best performing of which are thermally evaporated thin films of 3,4,9,10-perylenetetracarboxylic bis-benzimidazole (PTCBI). The initial results show device durability over 105 program/erase cycles, with hysteresis window of up to 3.3V, corresponding to charge storage density as high as 5×10^{12} cm^{-2}.

Data shows that charge retention is improved for molecular films with lower carrier mobility, which for the first time experimentally confirms in a coherent material set that inhibiting charge transport by nano-segmented floating-gate structures benefits the memory retention. These results show a first step towards a possible approach to miniaturization of non-volatile memory by using molecules as segmented charge storage elements in the floating gate flash memory technology.

Thesis Supervisor: Vladimir Bulović
Title: Professor

To my family

Acknowledgments

This work would not have been possible without the support of many talented people and organizations.
I would like to thank my research advisor Prof. Vladimir Bulovic for his guidance throughout the completion of this thesis. I feel very fortunate for the opportunity to have him as my research advisor.

Also, I would like to thank my colleagues who have taught me so much. Specifically, I would like to thank Dr. Ivan Nausieda, Dr. Polina Anikeeva, Dr. Osama Nayfeh, Hassen Abdu, Dr. Corinne Packard, and Prof. Dimitri Antoniadis for their assistance in completing of this manuscript.

Funding of this thesis was provided by the Focus Center Research Program Center for Materials Structures and Devices (MSD).

Contents

1 Introduction to Flash Memory 7

1.1 Memory Industry and Applications 7
1.2 Flash Memory Structure 8
1.3 Program and erase mechanisms 11
1.3.1 Hot-Electron Injection 11
1.3.2 Fowler-Nordheim Tunneling 12
1.4 Flash memory Reliability 13
1.4.1 Charge Retention 13
1.4.2 Endurance 14
1.5 Types of Flash Memory 15
1.6 Future Scaling Challenges 17
1.6.1 Physical cell-scaling challenges 18
1.6.2 Electrical cell-scaling challenges 19
1.6.3 Reliability scaling challenges 23
1.7 Summary 25

2 Molecular Floating Gate Memory 27

2.1 Segmented Floating Gate memories 27
2.2 Organic Molecular Floating gate Memories 29
2.2.1 Molecular Floating Gate Capacitors 31
2.2.2 Fabrication of MOS devices with Molecular Floating Gate 32
2.2.3 Results and Discussion 33

3 Conclusion 41

4 Bibliography 42

List of Figures

Figure 1-1:(a) Schematic cross section of conventional flash memory, (b) flash memory energy band diagram. 10
Figure 1-2: Schematic cross section and energy band diagram of conventional flash memory during programming and erasing. 11
Figure 1-3: Writing mechanism in floating-gate devices [14]. 13
Figure 1-4: Threshold voltage window closure as a function of program/erase cycles on a single cell [14]. 14
Figure 1-5: Schematic cross-sections and circuit diagrams for NOR and NAND flash memory. In NOR memory, the basic unit is one memory transistor. For NAND memory, the basic unit is 16 memory and 2 select transistors [17]. [WL: wordline; SG(D): select gate drain; SG(S). 16
Figure 1-6: Flash memory device scaling history (source: Dr. A. Fazio, Intel Corp.). 17
Figure 1-7:Schematic of a Flash array, showing row and column disturbs occurring when the cycled cell is programmed. 20
Figure 1-8: Minimization of crosstalk by replacing the floating gate with either floating traps or floating conducting islands. 20
Figure 1-9: High-level depiction of floating-gate transistor improvement. The ONO dielectric in the traditional transistor (left) is replaced by a high-k insulating dielectric (right). A floating-gate (poly in the diagram) is replaced by either floating. 21
Figure 1-10:Conduction band edge diagrams of typical uniform barrier and crested symmetric barrier 22
Figure 1-11:Threshold voltage distribution for 2 b/cell. 22
Figure 1-12:Number of Electrons Stored in a Floating Gate (source: Chung Lam, IBM). 25
Figure2-1: Segmented floating gate with 10-15 nm in diameter nanocrystals (source:freescale). 28
Figure 2-2: Charge loss through the oxide defects in memories with continues floating gate and segmented floating gate. 28
Figure 2-3: A 200 Å × 200 Å STM image of PTCDA on HOPG showing a unit cell consisting of two molecules indicating the crystallographic directions. Image taken in the constant current mode under UHV conditions, with a tip current of 200 nA and voltage of -800 mV [43]. 30
Figure 2-4: (a) Suggested energy-band diagrams of four different organic memory devices with chemical structure of each material (energy values are in eV). (b) The schematic cross-section of the device structure. 31
Figure 2-5: AFM (a) topographical and (b) phase images of 30nm-thick PTCBI layer. 33
Figure 2-6: The *C-V* characteristics of memory devices with a 10nm thick layer of PTCDA. 34
Figure 2-7: The *C-V* characteristics of memory devices with a 10nm thick layer of PTCBI. 34
Figure 2-8: Perspective views of a PTCBI and PTCDA unit cell [32]. 35
Figure 2-9: The *C-V* characteristics of memory devices with a 10nm thick layer of C60[44]. 36
Figure 2-10: The *C-V* characteristics of memory devices with a 10nm thick layer of Alq_3[44]. 36
Figure 2-11: The *C-V* characteristics of a control device without organic layer measured in the -8 V/8 V sweep range [44] 37
Figure 2-12: P/E endurance characteristics of memory devices with (a) PTCDA, (b) PTCBI, (c) C_{60} and (d) Alq_3 floating gates [44]. 38
Figure 2-13: Normalized retention characteristics measured at room temperature [44]. 39

List of Tables

TABLE I: QD-TO-QD DISTANCE FOR TECHNOLOGY DESIGN NODES[13] 29
TABLE II: 20% CHARGE LOSS TIME VERSUS ELECTRON LATERAL MOBILITY 40

Chapter 1

Introduction to Flash Memory

1.1 Memory Industry and Applications

Complementary metal-oxide-semiconductor (CMOS) memories can be divided into two main categories: volatile memories, which they lose stored information once the power supply is switched off, and nonvolatile memories which they keep stored information also when the power supply is switched off. In the past decade, memory chips with low power consumption and low cost have attracted more and more attention due to the booming market of portable electronic devices such as cellular phones and digital cameras. These applications require the memory to have ten years data retention time, so that the nonvolatile memory device has become indispensable. There are mainly four types of nonvolatile memory technology: flash memory, Ferro-electric Random Access Memory (FeRAM), Magnetic Random Access Memory (MRAM) and phase change memory. Flash memory is presently the most suitable choice for nonvolatile applications[2]. The continuous-film polysilicon-based floating-gate device has been the backbone of the nonvolatile memory (flash) market for the past decade.

The flash memory business flourished when the memory was adopted as the standard memory in cell phones, in which the memory enabled just-in-time loading of the latest program code as the last step in manufacturing, and program bugs could be fixed without taking the phone apart.

The simplicity of its device fabrication process is evident [1]. Flash memory fabrication process is compatible with the current CMOS process and is a suitable solution for embedded memory applications.

A flash memory cell is simply a MOSFET cell, except that a poly-silicon floating gate (or Silicon Nitride charge trap layer) is sandwiched between a tunnel oxide and an inter-poly oxide to form a charge storage layer. All other nonvolatile memories require integration

of new materials that are not as compatible with a conventional CMOS process. It is easier and more reliable to integrate flash memory than other nonvolatile memories with logic and analog devices in order to achieve better chip performance for wireless communication and wireless computation.

Flash memory can achieve the highest chip density since flash memory cell consists of only one transistor [3]. A FeRAM memory cell generally consists of one transistor and one capacitor [4], while a MRAM cell needs a transistor and a magnetic tunnel junction [5]. Phase change memory was expected to be a promising nonvolatile memory [6]; however, its memory cell consists of one resistor and a bipolar junction transistor. In addition, Flash memory possesses the multi-bit per cell storage property [7]. Four distinct threshold voltage (VT) states can be achieved in a flash memory cell by controlling the amount of charge stored in its floating gate[2].

However, after years of intense growth in the Flash memory market, conventional flash memory technology appears to be reaching fundamental scaling limits [1], [8]. The difficulty in scaling the tunnel oxide thickness due to leakage-current-related charge loss, reduction in gate coupling, and increase in cell-to-cell interference, necessitate modification in the design of the flash memory structures, including proposals for replacing the polysilicon floating gate by either floating traps such as silicon nitride in the SONOS technology [3], [4] or floating quantum dots (QD) [5]–[9]

1.2 Flash Memory Structure

To have a memory cell that can commute from one state to the other and that can store the information independently of external conditions, the storing element needs to be a device whose conductivity can be changed in a nondestructive way.

One solution is to have a transistor with a threshold voltage that can change repetitively from a high to a low state, corresponding to the two states of the memory cell, i.e., the binary values ("1" and "0") of the stored bit. Cells can be "written" into either state "1" or "0" by either "programming" or "erasing" methods. One of the two states is called "programmed," the other "erased."

The threshold voltage of a MOS transistor can be written as [1]

$$V_T = K - \bar{Q}/C_{ox}$$

where K is a constant that depends on the gate and substrate material, doping, and gate oxide thickness, Q is the charge weighted with respect to its position in the gate oxide, and C_{ox} is the gate oxide capacitance. As can be seen, the threshold voltage of the memory cell can be altered by changing the amount of charge present between the gate and the channel. There are many ways to obtain the threshold voltage shift. Two are the most common solutions used to store charge:

- In a conductive material layer between the gate and the channel and completely surrounded by insulator. This is the floating gate (FG) device.
- In traps that are present in the oxide, more precisely at the interface between two dielectric materials. The most commonly used interface is the silicon oxide/nitride interface. Devices obtained in this way are called metal-nitride-oxide-silicon (MNOS) cells [9], [10].

The schematic cross section of a generic FG device is shown in Figure 1-1(a); the upper gate is the control gate (CG) and the lower gate, completely isolated within the gate dielectric, is the FG. The FG acts as a potential well (see Figure 1-1(b)). If a charge is forced into the well, it cannot move from there without applying an external force: the FG stores charge [1].

Usually the gate dielectric, i.e., the one between the transistor channel and the FG, is an oxide in the range of 9–10 nm and is called "tunnel oxide" since electron tunneling occurs through it. The dielectric that separates the FG from the CG is formed by a triple layer of oxide–nitride–oxide (ONO). The ONO thickness is in the range of 15–20 nm of equivalent oxide thickness. The ONO layer as interpoly dielectric has been introduced in order to improve the tunnel oxide quality.

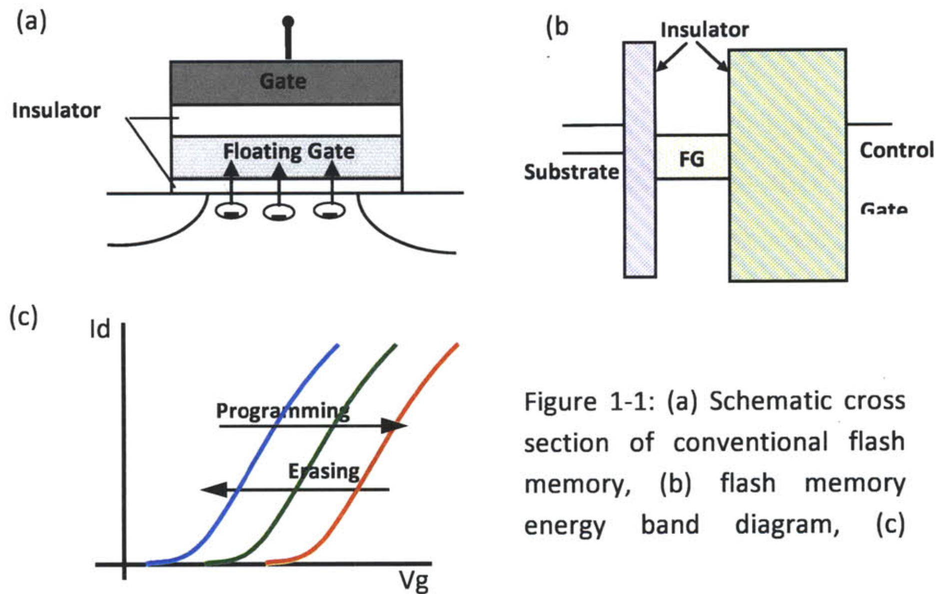


Figure1-1:(a) Schematic cross section of conventional flash memory, (b) flash memory energy band diagram

Stored charges on the floating-gate alter the threshold voltage of the MOSFET. The amount of charge stored on the floating gate can be controlled with biasing the terminal electrodes with voltages sufficient enough to cause tunneling of carriers in the gate-insulator from either the channel or the gate. These charges can tunnel through the gate insulator and be trapped/stored on the floating gate (programming), therefore causing a shift in the threshold voltage of the device. Appropriate biasing of the device can also be performed to cause removal of the charges that are stored on the floating gate (erasing), and to return the threshold voltage to the original uncharged state (Figure 1-2).

The data stored in a Flash cell can be determined measuring the threshold voltage of the FG MOS transistor by reading the current driven by the cell at a fixed gate bias.

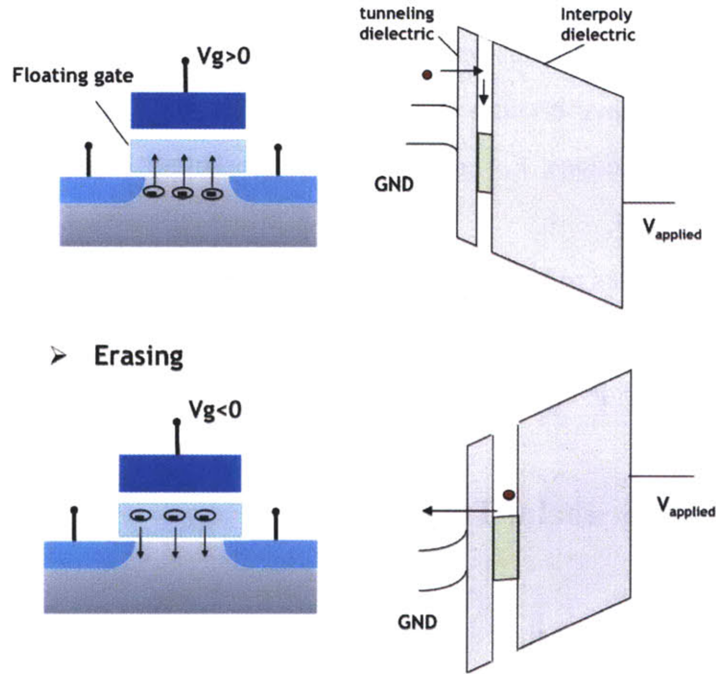


Figure 1-2: Schematic cross section and energy band diagram of conventional flash memory during programming and erasing.

1.3 Program and erase mechanisms

There are two main mechanisms by which charge carriers can charge and discharge the floating gate: Hot-Electron Injection and Fowler-Nordheim tunneling.

1.3.1 Hot-Electron Injection

Hot-Electron Injection is a phenomenon by which a charge carrier in the channel gains energy from the lateral electric field, and then crosses the oxide energy barrier into the floating gate by experiencing a vertical electric field between the control gate and substrate.

An electron traveling from the source to the drain gains energy from the lateral electric field and loses energy to the lattice vibrations (acoustic and optical phonons). At low fields, this is a dynamic equilibrium condition, which holds until the field strength reaches approximately 100 kV/cm [1],[11]. For fields exceeding this value, electrons are

no longer in equilibrium with the lattice, and their energy relative to the conduction band edge begins to increase. Electrons are "heated" by the high lateral electric field, and a small fraction of them have enough energy to surmount the barrier between oxide and silicon conduction band edges. For an electron to overcome this potential barrier, three conditions must hold [12].

1) Its kinetic energy has to be higher than the potential barrier.
2) It must be directed toward the barrier.
3) The field in the oxide should be collecting it.

1.3.2 Fowler-Nordheim Tunneling

Fowler-Nordheim Tunneling is another mechanism for charge carriers to cross the oxide energy barrier by applying a strong electric field (in the range of 8–10 MV/cm) across a thin oxide. The concept of tunneling is rooted in quantum mechanics. Electrons 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. For this method only single external power supply is needed [13].

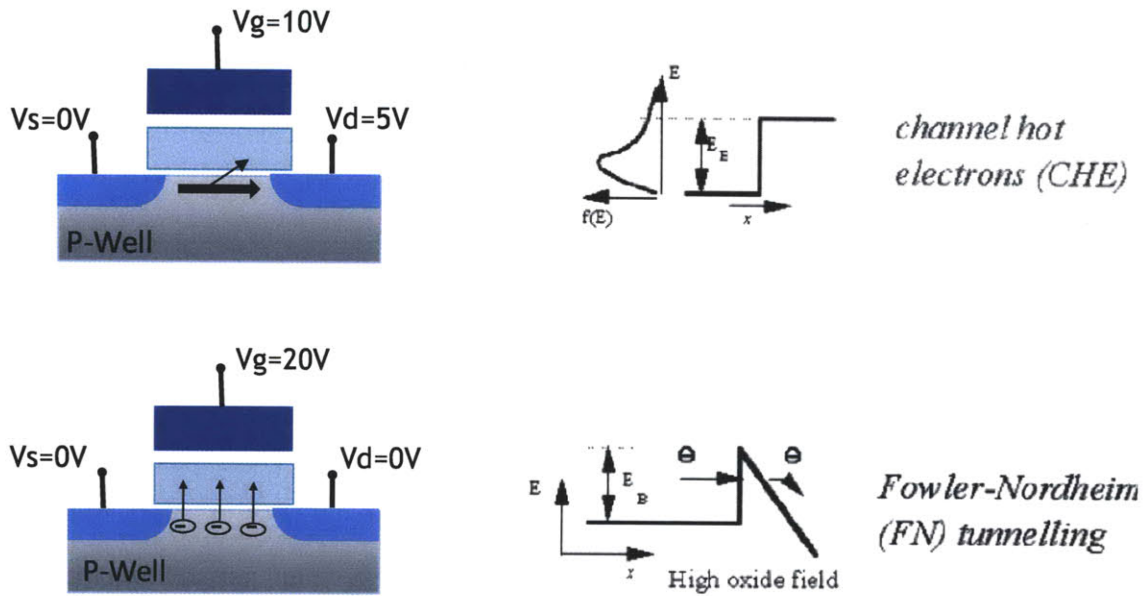


Figure 1 -3: Writing mechanism in floating-gate devices [14]

1.4 Flash memory Reliability

Reliability of a Flash cell is attributed to its endurance and retention. Endurance (capability of maintaining the stored information after erase/program/read cycling) and retention (capability of keeping the stored information in time) are the two parameters that describe how "good" and reliable a cell is.

1.4.1 Charge Retention

As in any nonvolatile memory technology, Flash memories are specified to retain data for over ten years. This means the loss of charge stored in the FG must be as minimal as possible.

Possible causes of charge loss are:

1) defects in the tunnel oxide;
2) defects in the interpoly dielectric;
3) mobile ion contamination; and
4) detrapping of charge from insulating layers surrounding the FG.

The generation of defects in the tunnel oxide can be divided into an extrinsic and an

intrinsic one. The former is due to defects in the device structure; the latter is due to the physical mechanisms that are used to program and erase the cell. The tunnel oxidation technology as well as the Flash cell architecture is a key factor for mastering a reliable Flash technology.

The best interpoly dielectric considering both intrinsic properties and process integration issues has been demonstrated to be a triple layer composed of ONO. For several generations, all Flash technologies have used ONO as their interpoly dielectric.

Electrons can be trapped in the insulating layers surrounding the floating gate during wafer processing, as a result of plasma damage, or even during the UV exposure normally used to bring the cell in a well-defined state at the end of the process. The electrons can subsequently detrap with time, especially at high temperature. This apparent charge loss disappears if the process ends with a thermal treatment able to remove the trapped charge.

The retention capability of Flash memories has to be checked by using accelerated tests that usually adopt screening electric fields and hostile environments at high temperature [14].

1.4.2 Endurance

A floating gate needs to endure over 10^5 program/erase cycles. Cycling is known to cause a fairly uniform wear-out of the cell performance, mainly due to tunnel oxide degradation, which eventually limits the endurance characteristics [15].

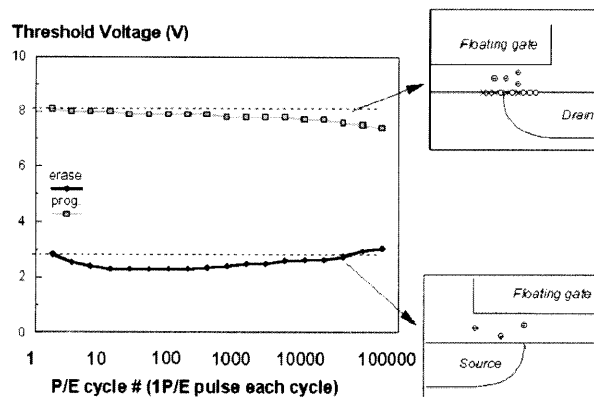


Figure 1 -4: Threshold voltage window closure as a function of program/erase cycles on a single cell [14].

A typical result of an endurance test on a single cell is shown in Figure 1-4. The reduction of the programmed threshold with cycling is due to trap generation in the oxide and to interface state generation at the drain side of the channel, which are mechanisms specific to hot-electron degradation. The evolution of the erase threshold voltage reflects the dynamics of net fixed charge in the tunnel oxide as a function of the injected charge. The initial lowering of the erase is due to a pile-up of positive charge which enhances tunneling efficiency, while the long-term increase of the erase is due to a generation of negative traps. Cycling wear-out can be reduced by proper device engineering and by optimization of the tunnel oxide process.

1.5 Types of Flash Memory

Two major forms of Flash memory, NAND Flash and NOR Flash, have emerged as the dominant varieties of non-volatile semiconductor memories utilized in portable electronics devices. NAND Flash, which was designed with a very small cell size to enable a low cost-per-bit of stored data, has been used primarily as a high-density data storage medium for consumer devices such as digital still cameras and USB solid-state disk drives. NOR Flash has typically been used for code storage and direct execution in portable electronics devices, such as cellular phones and PDAs.

In NOR flash memory, each cell resembles a standard MOSFET, except that the cell has two gates, stacked vertically, instead of just one. Each NOR memory cell is connected to the common drain connection called a bitline and can be read from directly giving the fast read performance that is necessary for fast program execution. In order to decrease the cost of flash memory, NAND flash memory (Figure 1-5) was invented [16],[17].

In NAND flash memory, the memory cells are connected in series with 16 or 32 memory cells connected to the bitline and source line through two select transistors. [In Fig. 5, the source line is connected to the ground through SG(S).] Because cell contact

area represents about 30% of unit cell area, this serial cell approach gives smaller cell size and lower die cost compared to NOR memory. The tradeoff is slower read performance because the read current is lower when using serial transistors. The NAND memory business flourished with the growth in popularity of digital cameras, for which NAND memory cards provided a convenient low-cost media for

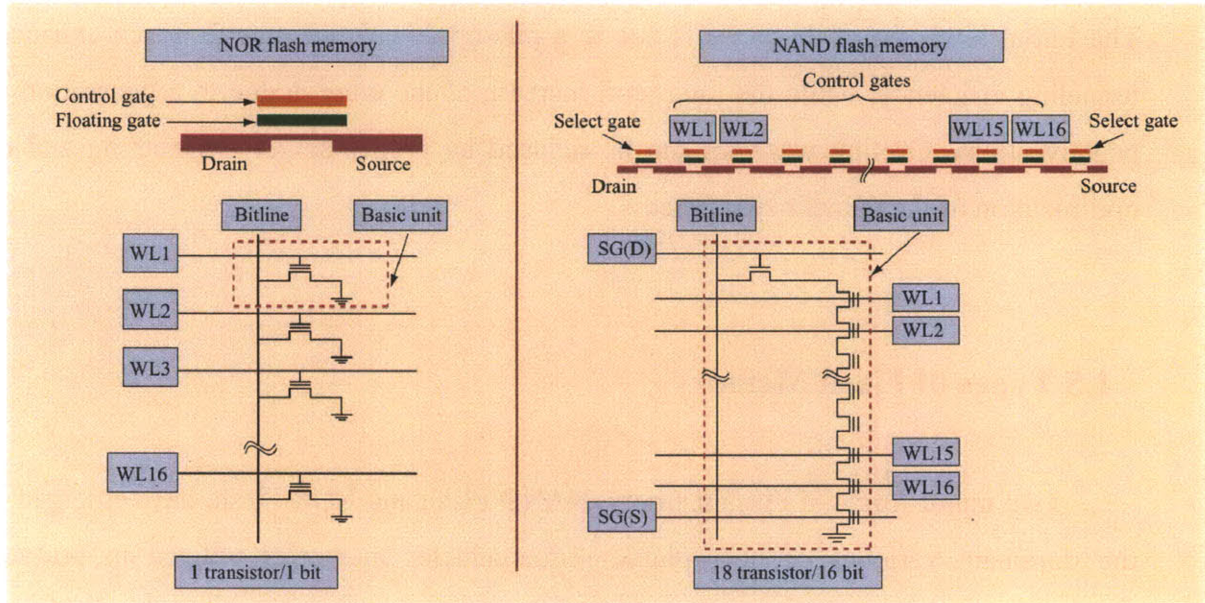


Figure 1-5: Schematic cross-sections and circuit diagrams for NOR and NAND flash memory. In NOR memory, the basic unit is one memory transistor. For NAND memory, the basic unit is 16 memory and 2 select transistors [17]. [WL: wordline; SG(D): select gate drain; SG(S)

picture storage. The slow read speed is not an issue for such applications. This application was followed by the ubiquitous USB (Universal Serial Bus) drives and MP3 (MPEG-1 Audio Layer 3) players. An emerging new application that will drive more growth for NAND memory is solid-state disks to replace disk drives in notebook computers. The growth of flash memory over the years was driven by the relentless memory cost reduction through Moore's Law; the price for flash memory dropped from approximately $80,000 per gigabyte in 1987 for NOR flash to approximately $10 per gigabyte in 2007 for NAND flash.

1.6 Future Scaling Challenges

NAND and NOR flash memories have had great success with respect to memory cell-size reduction and the corresponding product cost reduction. Looking toward the future, we expect significant scaling challenges. In most cases, even though innovations will exist to facilitate scaling, increasingly, they will involve a significant increase in complexity or the use of expensive new manufacturing tools. Thus, the scaling limit in the future may depend more on economics than on purely technical issues [17].

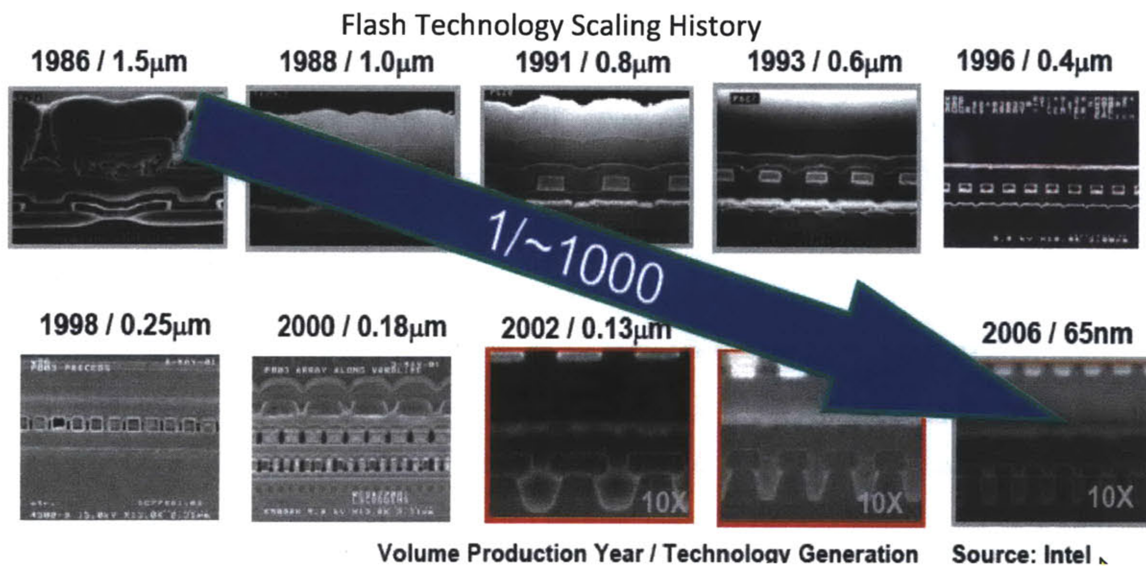


Figure 1-6: Flash memory device scaling history (source: Dr. A. Fazio, Intel Corp.).

Broadly speaking, there are three key areas of challenges for flash memory scaling: 1) physical scaling, which is primarily defined by lithography and the cell layout design; 2) electrical scaling, which is primarily defined by the program/erase/read voltage requirements; and 3) reliability scaling, which is primarily defined by the fundamental physics of the program, erase, and storage mechanisms.

1.6.1 Physical cell-scaling challenges

In all cases, lithography is still the main factor affecting Moore's Law scaling. NAND memory, with its very regular layout consisting of straight lines and spaces, allows for the use of many optical enhancement technologies, which greatly facilitate the continued use of conventional optical lithography. Consequently, NAND memory leads the industry with the tightest lithographic pitch of any silicon memory products. However, conventional lithography with i-line and immersion technology is affected by a manufacturing limit at approximately 40 nm. To go beyond conventional lithography, new techniques such as self-aligned double patterning have been reported [18]. By using spacers on two sides of the lithographic line, the space between lines can be further subdivided, effectively improving the resolution. Given this feature, the ability to define minimum line and space is extended to nearly 20 nm and will mostly likely not be the limiting factor for scaling. Note that this technique adds exposure and other process steps that increase the cost of pattern definition. In the case of NOR memory, as far as the 65-nm generation, the layout of the memory cell has 45-degree angle structures around source contacts that are not conducive to optical enhancement techniques [19]. To improve NOR scaling, a self-aligned contact technology is being developed for the 45-nm lithographic node [19], which gives two significant improvements. First, the self-aligned contact reduces the contact area, a scaling limiter for NOR flash memory. Second, the new layout consists of straight lines only similar to NAND, making it easier to implement optical enhancement techniques.

To summarize, in both NOR and NAND memories, it is possible to continue to reduce the physical size of the cell to dimensions close to 20 nm. The true limiters of cell-size reduction involve electrical and reliability requirements, which is discussed in the following sections [17].

1.6.2 Electrical cell-scaling challenges

For NOR flash memories, a primary scaling limitation for the cell is the high voltage required during the programming operations, which in turn limits the minimum channel length. To achieve hot carrier channel programming, a voltage of more than 4 V is required from the drain to the source to produce electrons of sufficient energy to overcome the 3.2-eV Si-to-SiO2 barrier height [20]. Therefore, the minimum gate length will be limited to the channel length that can withstand the required programming voltage. For NAND flash, the transistor channel is used for read only, requiring a much lower drain-to-source voltage and, therefore, a shorter channel length limit.

Three-dimensional cell structures are one way to address the gate length scaling constraint. Both above- silicon fin structures [20], [21] and below-silicon U-shaped structures [22] have been reported. These structures move the channel length constraint into the Z direction, allowing further X/Y scaling to occur. This permits further area scaling while maintaining the total channel length required. Recent experimental results reported for NAND, involving the hemi-cylindrical FET (HCFET) [18], show superior transistor characteristics down to the 38-nm node.

Another significant scaling limitation involves maintaining adequate coupling of the control gate to the floating gate. A high coupling ratio is required to provide adequate control of the channel used for reads. As the cell scales in size and self-aligned techniques are used for the floating gate, maintaining control of the channel requires a thinner inter-poly dielectric between the control gate and the floating gate. One possible solution is the use of a high-k dielectric (i.e., a dielectric material with a high dielectric constant). It must be emphasized that the requirement of a high-k dielectric for the inter-polysilicon layer is different from a high-k dielectric for the transistor gate. The inter-poly dielectric has to be optimized for no leakage current under low-field charge storage, whereas a gate dielectric can have a small leakage current.

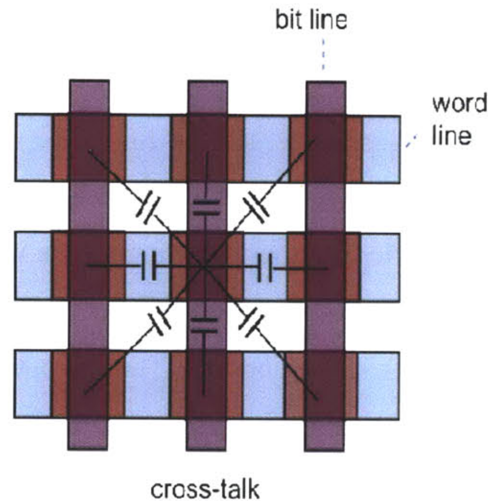


Figure 1-7:Schematic of a Flash array, showing row and column disturbs occurring when the cycled cell is programmed.

Another scaling limitation involves the coupling of two adjacent cells through the capacitance between the cells [20],[22]. As the spacing between floating gates is reduced, the floating-gate to floating-gate coupling increases. The data stored in one cell can influence the operation of an adjacent cell. Different solutions exist to address this problem, including reducing the size of the floating gate, electrical screening of the floating gate, or special read biases to compensate for the coupling. In the case of NAND memory, the most promising approach is to replace the floating gate with either floating traps [25] or floating conducting islands that function as charge storage layers. In such cases, the capacitive coupling between adjacent cell charge storage layers is greatly reduced (Figure 1-8). However, this is not a possible solution for floating-gate NOR memory, because in order to move across the transistor channel,

12

Figure 1-8:Minimization of crosstalk by replacing the floating gate with either floating traps or floating conducting islands

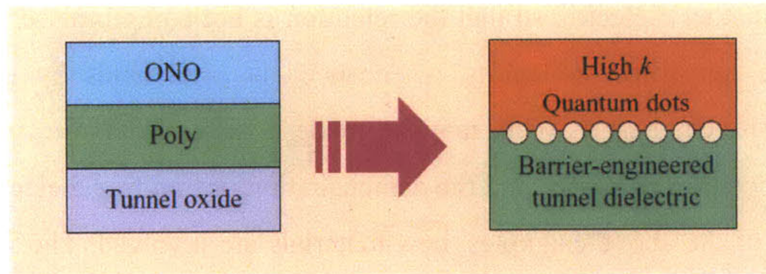


Figure 1.9: High-level depiction of floating-gate transistor improvement. The ONO dielectric in the traditional transistor (left) is replaced by a high-k insulating dielectric (right). A floating-gate (poly in the diagram) is replaced by either floating

NOR memory relies on a conducting charge storage layer to redistribute the channel hot electron charge injected in the drain area.

To summarize, the general concept of extending scaling limit is shown in Figure 1-9. First, ONO (oxide-nitride- oxide) scaling can be extended by the use of a high-k dielectric. Second, the polycrystalline floating gate can be replaced by either floating traps or floating quantum dots. Last, alternative materials can be explored to allow further improvement of the tunneling dielectric. Tunnel oxide of non-volatile memory (NVM) devices would be very difficult to downscale if ten-year data retention were still needed. This requirement limits further improvement of device performance in terms of programming speed and operating voltages.

When tunnel oxide for non-volatile memory (NVM) is scaled, the direct tunneling effect becomes dominant. A simple calculation shows that the minimum thickness is around 6 nm. In addition, the strain-induced leakage current increases for thinner oxides and aggravates scaling of tunnel oxide thickness. In current flash NVM devices, the tunnel oxide thickness is approximately in the 7–8 nm range. When the barrier is thin, the program and erase process is rapider but charge leakage destroys the retention time. When the barrier is relatively thick, long charge retention times are achieved but a higher voltage and a longer period of time are required to program and erase the floating gate. Because of this, downscaling of tunnel oxide would be very difficult if ten-year data retention were still required. Consequently, for low-power applications with Fowler-Nordheim (FN) programming, such as NAND, program and erase voltages are essentially sustained at unacceptably high levels. A promising solution for tunnel oxide scaling is

engineering the tunnel dielectric so that the retention is not compromised at low electric fields while the tunneling probability is enhanced at high fields by using multiple dielectrics of different barrier. This may be made possible with the improvement in atomic-layer deposition processes and the demonstration of trap-free dielectric films [24]. Note that in all of the above examples, new materials are involved. The introduction of new materials in semiconductor manufacturing is a key part of the innovations that enabled Moore's Law scaling to continue for so many generations.

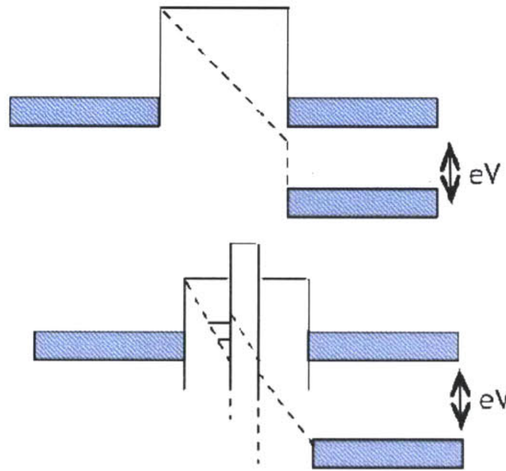


Figure 1-10Conduction band edge diagrams of typical uniform barrier and crested symmetric barrier

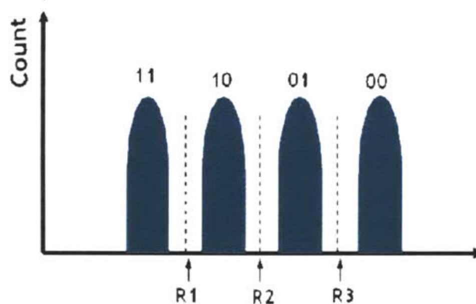


Figure 1-11Threshold voltage distribution for 2 b/cell

1.6.3 Reliability scaling challenges

One of the most significant innovations in both NOR and NAND flash memories is multilevel cell (MLC) technology: the storage of more than one bit in a single flash cell [26]. This is possible for flash memory because of the analog nature of charge storage in the floating gate, which allows for subdividing the amount of stored charge into small increments. When this is coupled with the superior retention characteristics of the floating gate, it is possible to accurately determine the charge state after a long period of time. The weakness of MLC arises because the separation between charge states is less for MLC technologies compared to SLC (single-level cell) technologies, resulting in a higher sensitivity to cell degradation mechanisms. To achieve stable storage, it is important to properly control the write and erase operations, using special MLC charge-placement algorithms, to reduce the damage of the tunnel dielectric by reducing the applied fields and controlling how the fields are increased or decreased with the controller rate during write and erase. A further enhancement is the use of error-management techniques, such as error correction, which can recover data or prevent errors.

As the memory cell is scaled, the cell capacitance is decreased, resulting in the decrease of charge stored [20]. For NOR flash with a larger memory cell layout, the number of stored electrons is approximately 1,000 for the 45-nm node, while for NAND flash [22], it is less than 500. In this case, for two-bit-per-cell with four-level MLC technology, the number of electrons per level is just more than 100. While the numbers of stored electrons decrease with each new lithography node, the defect charge leakage mechanisms causing charge loss remain the same. Thus, the impact of each defect on the cell-threshold voltages is proportionally larger for each new node, manifesting as faster threshold voltage drops and an increase in error rates. One method of mitigation involves the improvement of the tunnel dielectric to make it more resistant to defect generation by the introduction of nitrogen into silicon dioxide. Another method is to replace floating gates with either floating traps such as silicon nitride or floating dots such as silicon islands or metal nanodots. With discrete charge storage, the impact of a defect is limited only to charge stored in its proximity and not to leaking of the conducting floating gate.

However, the discrete traps or islands may store a smaller number of electrons compared to a floating gate, further exacerbating the decrease in stored electrons for each storage level. Even though this problem exists in both NOR and NAND memories, for actual products, it is a larger challenge for NOR flash memories because for NOR flash used in program execution, data errors will result in system failure, and the fast-read requirement does not give much time for error detection and corrections. For NAND flash used in secondary data storage, it is possible to implement extensive error corrections through sophisticated data controllers. With the error rate increasing with scaling, it is possible to implement in the data controllers increasing sophisticated error correction techniques that have been developed for the disk drive industry, and which have made disk drivesone of the more reliable storage devices [17].

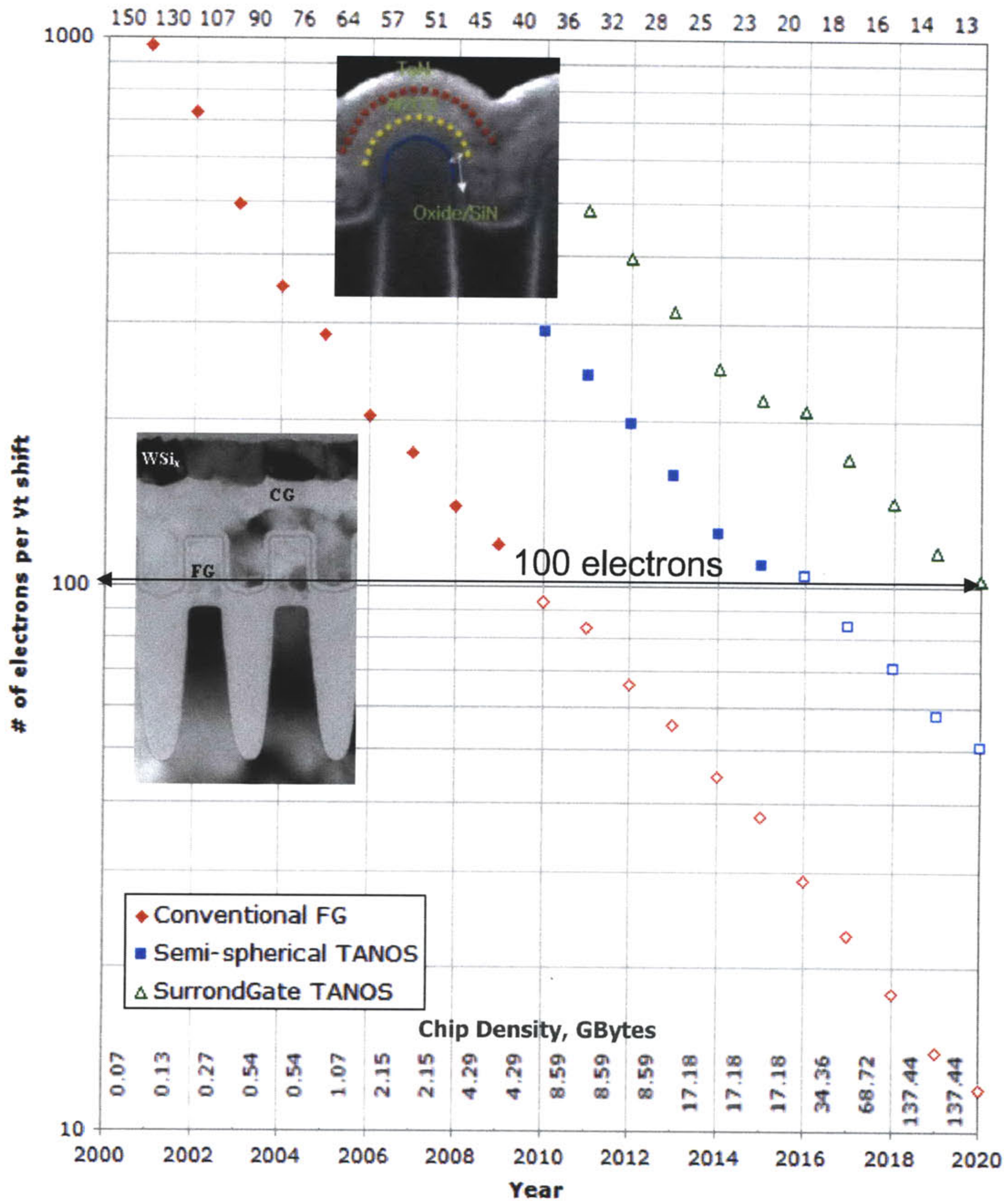


Figure 1-12Number of Electrons Stored in a Floating Gate (source: Chung Lam, IBM)

1.7 Summary

After years of intense growth in the Flash memory market, conventional flash memory technology appears to be reaching fundamental scaling limits [1], [27]. The difficulty in scaling the tunnel oxide thickness due to leakage-current-related charge loss, reduction in gate coupling, and increase in cell-to-cell interference, necessitate modification in the design of the flash memory structures, including proposals for replacing the polysilicon

floating gate by either floating traps such as silicon nitride in the SONOS technology [25], [28] or floating quantum dots (QD) [29], [30].

Chapter 2

Molecular Floating Gate Memory

2.1 Segmented Floating Gate memories

As demands for high storage density, high chip memory capacity, and decreasing process costs continue, conventional flash memory technology appears to be reaching fundamental scaling limits [1]. The minimum tunnel oxide thickness, already used in polysilicon flash memories[25], and poor charge retention due to defects in the tunneling oxide, necessitate modification in the implementation of the flash memory technology, including proposals for replacing the polysilicon floating gate with a sectioned gate consisting of multiple nanoparticle quantum dots (QDs) or floating traps such as silicon nitride in the SONOS technology [25], [28]. In Conventional flash memory, since the floating gate is conductive, the electrons can move freely in the conduction band and hence in case of any defect chain within the tunnel oxide, all of the trapped electrons in the floating gate can easily leak to the channel or source/drain through it. While the discrete charge storage in nano-segmented floating gate limits the impact of any tunnel oxide defects to the charge stored in the proximity of the defect site. Charge stored in the remaining parts of the nano-segmented floating gate remains unaffected due to the low charge mobility in the nano-segmented films, extending the functional retention time of the memory cells [31].

Another scaling limitation involves the coupling of two adjacent cells through the capacitance between the cells. As the spacing between floating gates is reduced, the floating-gate to floating-gate coupling increases. The data stored in one cell can influence the operation of an adjacent cell. Different solutions exist to address this problem, including reducing the size of the floating gate, electrical screening of the floating gate, or

special read biases to compensate for the coupling. However, in the case of NAND memory, replacing the floating gate with either floating traps or floating conducting islands is the most promising approach. In such cases, the capacitive coupling between adjacent cell charge storage layers is greatly reduced. However, this is not a possible solution for floating-gate NOR memory, because in order to move across the transistor channel, NOR memory relies on a conducting charge storage layer to redistribute the channel hot electron charge injected in the drain area [31].

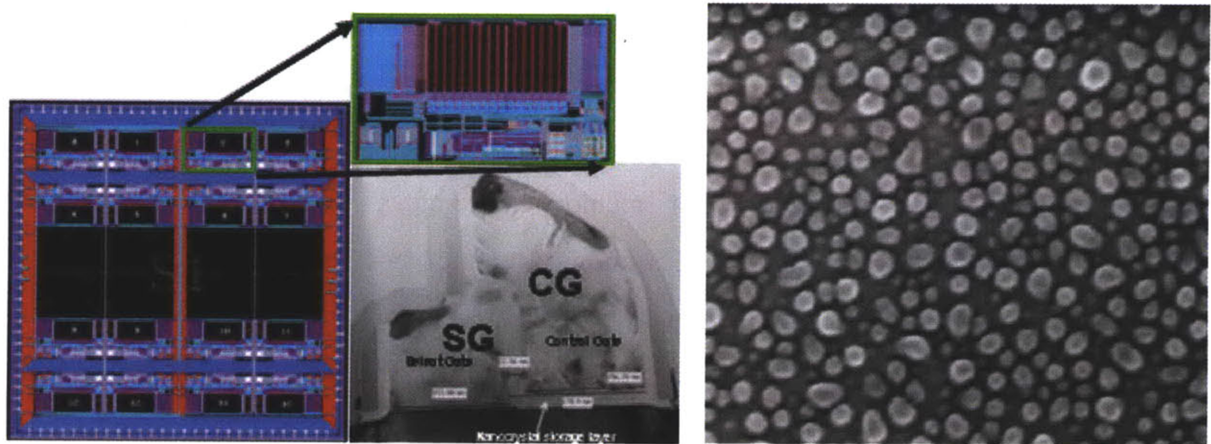


Figure2-1: Segmented floating gate with 10-15 nm in diameter nanocrystals (source:freescale)

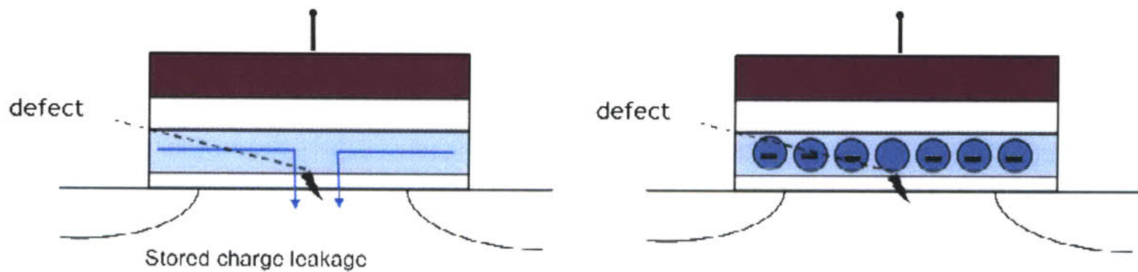


Figure 2-2: Charge loss through the oxide defects in memories with continues floating gate and segmented floating gate

In QD memories, for example, as shown in figure 2-2, the stored charge is distributed over multiple QD sites, so that if one of the QDs is discharged due to defects in the very thin tunneling oxide, the remaining QDs in the floating gate would retain their charge and preserve the memory state of the cell.

TABLE I
QD-TO-QD DISTANCE FOR TECHNOLOGY DESIGN NODES[13]

Technology Design Node	V = +/- 10%	V = +/-5%
50nm	7nm	4nm
39nm	6nm	3nm
28nm	4nm	2nm
20nm	3nm	1nm

Morphology experiments on QD monolayers together with numerical analysis of spatial packing of monodispersed nanoscale QDs indicate that use of the QD floating gate in the 50nm technology node would require use of hexagonally packed ordered QD monolayers with a QD-to-QD center distance spacing of 4nm or less [13]. Such closely spaced QDs could be susceptible to significant QD-to-QD charge tunneling, obviating the intended benefit of nanostructuring the gate electrode to preserve charge on individual QDs. This problem would be especially severe if metallic nanoparticles are utilized, as the large electron state densities and delocalized electron wavefunctions of metallic nanoparticles would facilitate efficient exchange of charges between closely packed QDs.

One solution to these problems is to utilize alternate charge storage elements, with small carrier state densities, and high charge carrier binding energies. Molecular materials that are on the order of 1nm in size, represent such idealized charge storage elements.

2.2 Organic Molecular Floating gate Memories

Molecular electron wavefunctions are strongly localized within the spatial extent of the molecules due to the charge binding energies that are on the order of 0.2 eV or higher [32] . A floating gate consisting of a thin film of molecules would provide the advantage of a uniform set of identical nanostructured charge storage elements with high molecular area densities (e.g. $8 \times 10^{13} cm^{-2}$ for PTCBI thin films used in this study) that can results in several-fold higher density of charge-storage sites as compared to QD memory and even SONOS devices [29].

Additionally, the low density of free carriers in the molecular thin films and the high

charge binding energy on individual molecules limit intermolecular interactions. The minimal overlap between the neighboring molecular electron wavefunctions contributes to the low organic thin film electron/hole mobilities, in the range of from 10^{-1} $cm^2/(Vs)$ to 10^{-7} $cm^2/(Vs)$, which contributes to the immunity of stored charge to the structural defects in the neighboring areas of the device.

Also a larger conduction band offset between the tunnel oxide and charge trap layer is desirable for mitigating trapped charge leakage into the substrate and consequently achieving longer retention time. Considering the lowest unoccupied molecular orbital (LUMO) of the utilized molecules shown in Figure 2-4(a), they provide fairly deep trapping sites compared to polysilicon in conventional flash memories or silicon nitride in trap based memories [32].

Finally, as compared to QDs, which typically exhibit size and order variability, molecular films have the highly desirable consistency of size and morphology, which provide relative constancy in the electronic energy level structure of molecular films. Although organic compounds have recently attracted growing interest for nonvolatile memory applications, many of the devices reported so far are two-terminal resistive memories [33]-[36], rather than reversible charge-storage elements, as described in this work.

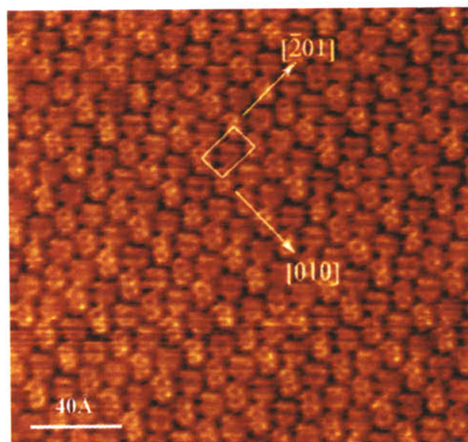


Figure 2-3: A 200 Å × 200 Å STM image of PTCDA on HOPG showing a unit cell consisting of two molecules indicating the crystallographic directions. Image taken in the constant current mode under UHV conditions, with a tip current of 200 nA and voltage of -800 mV [43]

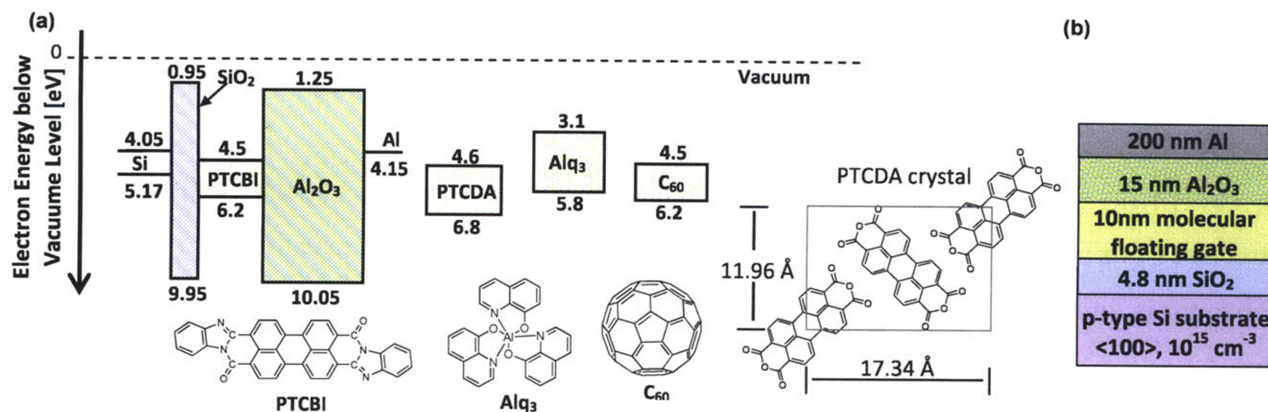


Figure 2-4: (a) Suggested energy-band diagrams of four different organic memory devices with chemical structure of each material (energy values are in eV). (b) The schematic cross-section of the device structure.

2.2.1 Molecular Floating Gate Capacitors

In this study we investigate charge-storage behavior in a series of molecular thin films embedded in metal-oxide-semiconductor (MOS) structures with SiO_2 and Al_2O_3 as the tunneling and control oxides, respectively. By comparing performance of different devices we identify the molecular thin film characteristics best suited for design of floating gate memories. We fabricate capacitive memory structures using archetypical molecular thin films with different charge storage energy levels and charge mobility including 3,4,9,10-perylenetetracarboxylic dianhydride (PTCDA), 3,4,9,10- perylenetetracarboxylic bis-benzimidazole (PTCBI), tris-(8-hydroxyquinoline) aluminum (Alq3), and fullerene (C_{60}).

The charge storage, retention, and program/erase endurance characteristics are examined via the capacitance voltage (*C-V*) measurements at frequency of 100 kHz, by using Agilent 4294 impedance analyzer at room temperature. The stored charge densities are determined by measuring the shift in the flat band voltage of molecular-film-containing capacitors. Data shows that charge retention times are improved for molecular films with lower carrier mobility, which for the first time confirms the stated operational benefit of the nano-segmented floating-gate structures, i.e. that lower charge mobility in the nano-segmented floating gate inhibits stored charge loss.

2.2.2 Fabrication of MOS devices with Molecular Floating Gate

The energy band diagrams and the schematic cross-section of the capacitive floating gate structures fabricated in this study are shown in Figure 2.4 The molecular-thin-film-containing capacitor is fabricated starting with a growth of a 4.8 nm thick layer of thermal SiO_2 (at 800 °C in dry O_2) on top of a cleaned p-type Si substrate.

Wafer cleaning was accomplished by immersing cassettes of wafers into cleaning baths containing SC-1 solution (1:1:5 $NH_4OH/H_2O_2/H_2O$) for striping organics, metals and particles, 50:1 H_2O/HF for removing the chemical oxide layers that are grown during the first step, and SC-2 solution (1:1:6 $HCL/H_2O_2/H_2O$) for striping alkali ions and metals. The wafers were rinsed with DI H_2O before and after the HF step and also after the last step.

The molecular material is thermally evaporated at the rate of (0.15 ± 0.05) nm/s at a base pressure of $<6\times10^{-7}$ Torr to form a 10 nm thick layer.

Prior to deposition, the organic materials were purified in three cycles using thermal gradient sublimation in a three-foot long Pyrex tube that was heated between 450°C and 100°C along its length, thereby allowing for the separation of impurities of both high and low volatility from the organic material. The pressure inside the purification tube was maintained at 10^{-2} Torr.

The surface morphology of the molecular layers was studied with atomic force microscopy (AFM). A representative AFM image, shown in Figure 2-5, of 30nm-thick PTCBI on 4nm-thick thermal SiO_2 on Si substrate, shows that PTCBI has poly crystalline morphology with average roughness of 1.42 nm.

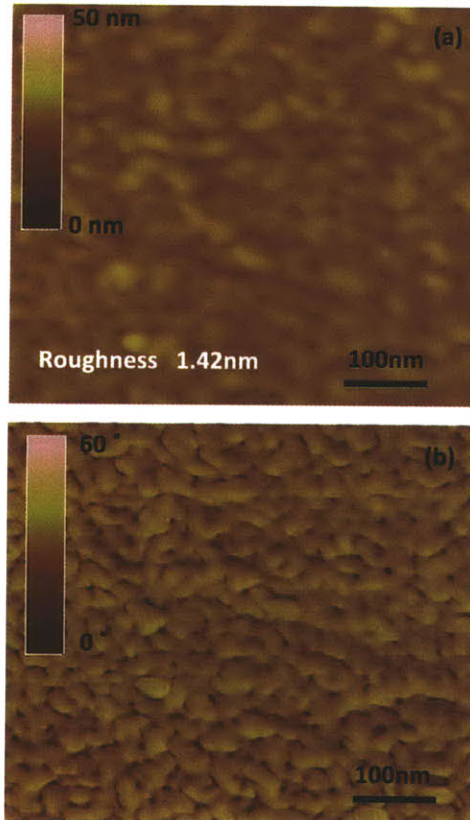


Figure 2-5: AFM (a) topographical and (b) phase images of 30nm-thick PTCBI layer

A 15 nm Al_2O_3 capping layer is deposited by RF magnetron sputtering on top of the organic layer, which is followed by deposition of a 300 nm thick Al film, which serves as a gate electrode. The processing is completed by annealing the devices in N_2 ambient at 275°C for 2.5 hours.

2.2.3 Results and Discussion

Figure 2-6 plots C-V characteristics of a device with a 10nm thick PTCDA layer as a floating gate. A clockwise hysteresis window of 2.5 V is observed for this device upon double sweeping within the range of from -10V to 10 V, with voltage held for 20sec at each bias. (In these measurements Si substrate is grounded, with the Al electrode biased as indicated.) The charge on the floating gate screens the applied electric field, which is manifested as an increase in the voltage needed to induce depletion and inversion in the Si semiconductor channel below. During

the negative sweep (from +10V to -10V) the floating gate is discharged and the C-V curve shifts back to the initial state. This memory device has a negligible hysteresis in the uncharged condition as the bias is swept forward and backward between -3V and 0V.

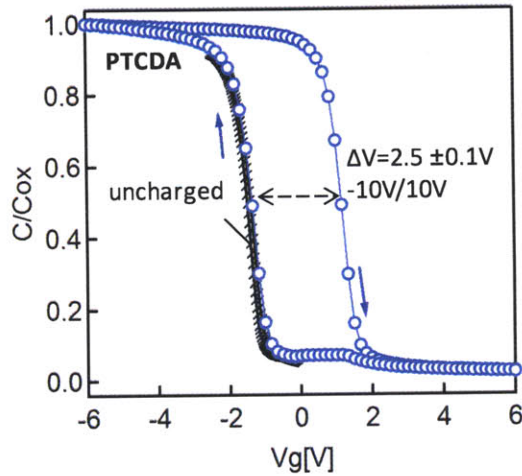


Figure 2-6: The *C-V* characteristics of memory devices with a 10nm thick layer of PTCDA

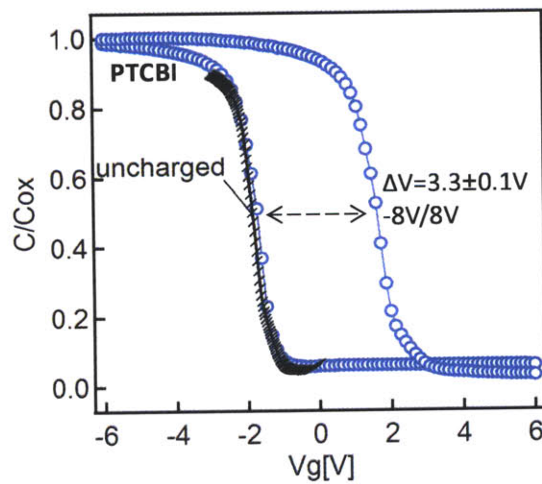


Figure 2-7: The *C-V* characteristics of memory devices with a 10nm thick layer of PTCBI

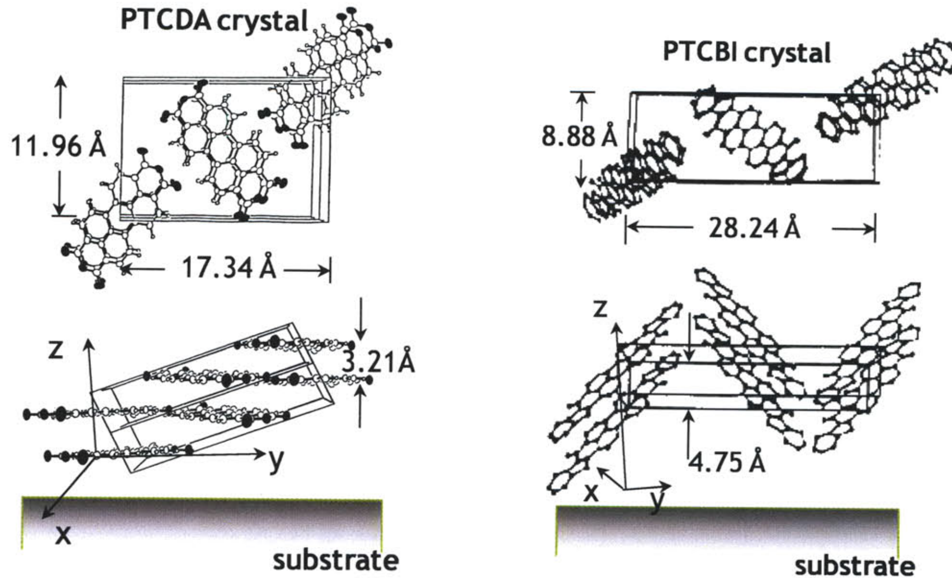


Figure 2-8: Perspective views of a PTCBI and PTCDA unit cell [32]

The device containing the PTCBI floating gate layer has a clockwise hysteresis, showing a rigid shift of (3.3±0.1) V in the C-V characteristics. The voltage shift to the right during the positive sweep (from -8V to +8V) is indicative of electron charging on the PTCBI floating gate (Figure 2-7).

PTCDA and PTCBI are planar molecules (see Figure 2-8) that form a herringbone molecular packing structure when crystallized [32],[37]. Unit cells of ordered monolayers of PTCDA and PTCBI contain two molecules each and occupy areas of 2.07 nm^2 and 2.51 nm^2, respectively, which correspond to molecular area densities of 9.7×10^{13} cm^{-2} and 8×10^{13} cm^{-2}, respectively [37]. From the flat band voltage shift we estimate [1] the stored charge density of 2.8×10^{12} cm^{-2} for PTCDA and 5×10^{12} cm^{-2} for PTCBI memory devices.

C_{60} memory devices showed a remarkably large hysteresis window of (6.0±0.1)V for program/erase condition of -8V/8V (Figure 2-9). A voltage shift to the right during the positive sweep (-8V to +8V) and a voltage shift to the left during the negative sweep (+8V to -8V) is indicative of both electron and hole charging on the C_{60} floating gate with estimated stored electron and hole densities of 3.6×10^{12} cm^{-2} and 2.9×10^{12} cm^{-2}, respectively. The large hysteresis window of C_{60}-film-containing capacitive structures can be explained by noting the high mobility of C_{60}, as compared to PTCDA, PTCBI, and Alq_3 [38]-[40] which leads to rapid

charge distribution across the floating gate during programming and facilitates ease of charging when the charge injection into the floating-gate is non-uniform across the floating-gate molecular film.

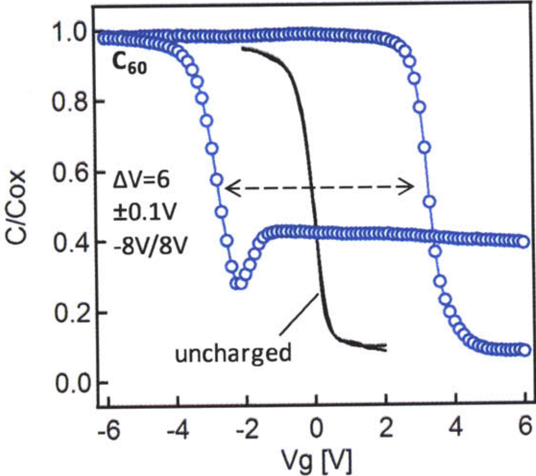


Figure 2-9: The *C-V* characteristics of memory devices with a 10nm thick layer of C60[44]

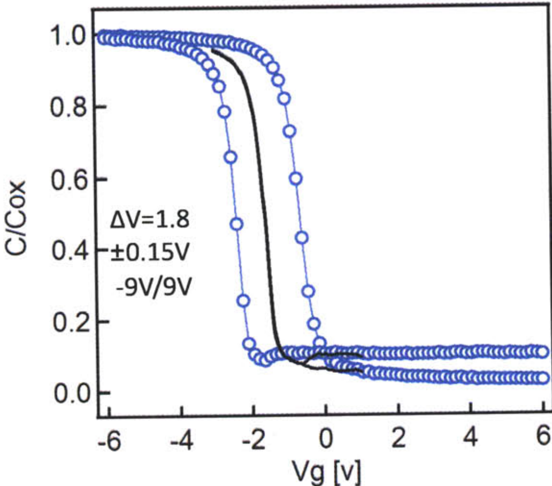


Figure 2-10: The *C-V* characteristics of memory devices with a 10nm thick layer of Alq_3[44]

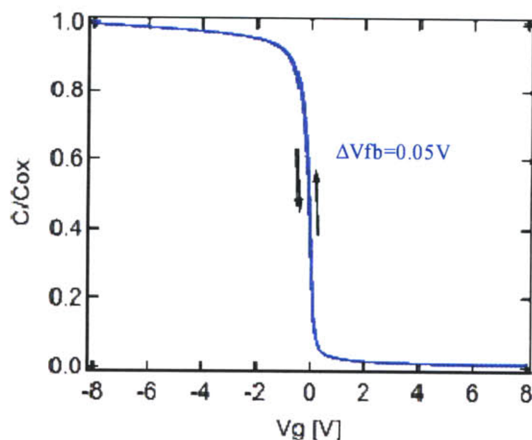


Figure 2-11: The *C-V* characteristics of a control device without organic layer measured in the -8 V/8 V sweep range [44]

Our measurements indicate that devices containing thin films of Alq_3 molecules show comparatively poor memory behavior compared to other tested materials. As shown in Figure 2-10 the Alq_3 device has a clockwise hysteresis window of (1.8±0.15) V for forward and backward sweep between -9Vand 9V.

In contrast to the molecular-film-containing capacitive memories, control devices that have no molecular layer show a minimal hysteresis of 0.05V for program/erase conditions of +8V/-8V (inset of Figure 2-11), which is consistent with charge storage in the molecular films.

The cycling endurance results for PTCDA and PTCBI memory devices are plotted in Figure 2-12 (a),(b), showing a +/-10% variation in the flat band voltage after more than 10^5 programming and erasing cycles. Over the same number of cycles the endurance characteristics of C_{60} samples exhibit only a slight increase in the flat band voltage shift (Figure 2-12 (c)), while Alq_3 devices show large change in the flat band voltage values during program/erase cycling (Figure 2-12(d)).

The structural disorder that leads to the low charge mobility in the amorphous Alq_3 molecular thin films can explain the observed gradual change in the flat band voltage with program/erase cycling. Based on the Monte Carlo calculations of Madigan and Bulović [42] the structural disorder in the amorphous thin films of polar Alq_3 molecules contributes to the energetic disorder in the electronic states of Alq_3, which is manifested as broadening of the distribution of the available electronic energy levels [42]. At the low-energy tail of this distribution are the deep electronic trap states that can be difficult to discharge [42]. The gradual increase in the density of

stored electrons in these deep traps during the successive program/erase cycles can be responsible for the observed flat band voltage shift in Alq_3 films towards the positive voltage.

Charge retention experiments are performed only on PTDA, PTCBI and C_{60} samples, as these structures demonstrate adequate cycling endurance for flash memory requirement and larger hysteresis windows compared to Alq_3 samples. Due to the leakage through control oxide film, the measured charge retention times for these memories are not sufficient to meet data retention standard for flash memories. Nevertheless, dependence of charge retention time on the charge mobility in the utilized molecular thin films is apparent.

Flatband voltage shifts at room temperature are plotted against the retention time in Fig. 5, with all the plots normalized to the initial shift of each device.

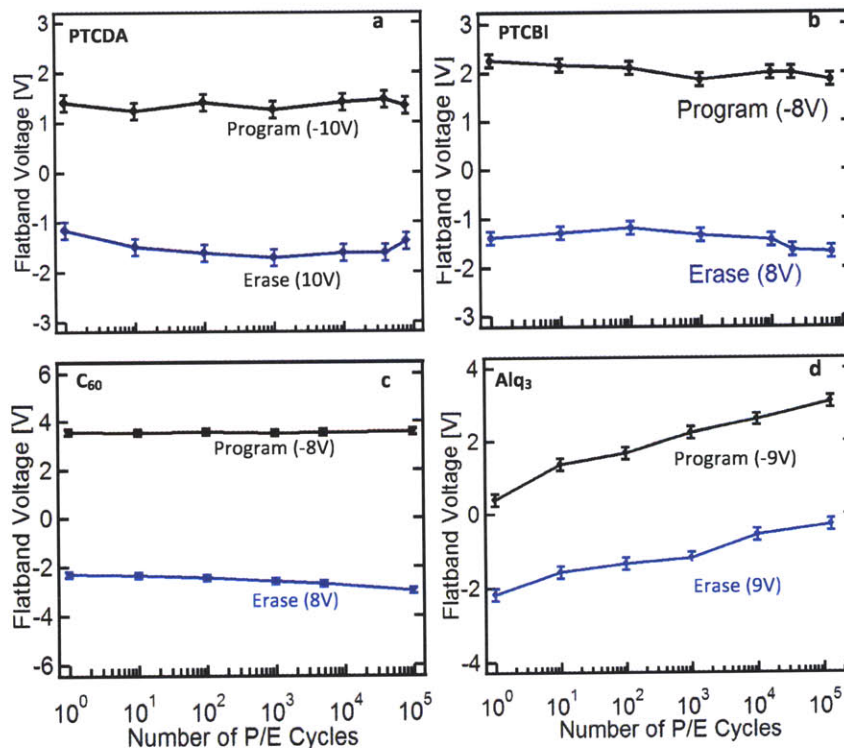


Figure 2-12: P/E endurance characteristics of memory devices with (a) PTCDA, (b) PTCBI, (c) C_{60} and (d) Alq_3 floating gates [44].

As shown in Figure 2-13 the retention loss of C_{60} sample (77% charge loss in 12 min) is found to be much larger than that of the PTCDA (programmed with -10V) and PTCBI (programmed

with -8V) samples which showed 20% charge loss after 11min and (1.5±0.2) hours, respectively.

The low retention time of C_{60} memory devices compromises their otherwise remarkable memory characteristics. We suggest that the low retention time of C_{60}-containing memory is due to high electron mobility (0.05 $cm^2V^{-1}s^{-1}$) of C_{60} which leads to lateral transport of electrons to the existing defects in oxide layers [38]. In contrast, PTCDA thin films behave as one-dimensional conductors, with charge transport confined to the molecular stacking direction, which is typically normal to the substrate surface. In the direction parallel to molecular planes, carrier mobilities range from 10^{-4} to 10^{-5} $cm^2V^{-1}s^{-1}$ [39].

Although PTCDA and PTCBI are very similar in their molecular structures; the larger interplanar stacking distance of PTCBI leads to considerably reduced π-orbital overlap and electronic anisotropy. The herringbone packing of PTCBI molecules separates the π-electron clouds on molecular neighbors, which results in the low electron mobility ($\mu = 2.4\times10^{-6}$ $cm^2V^{-1}s^{-1}$) of PTCBI thin films in the direction parallel to molecular planes [39], [42].

Such low charge mobility provides the same advantage as earlier demonstration of QDs in the floating gate memories, namely, if a defect exists in the tunneling oxide below the floating gate, charges in the floating gate are unlikely to transport laterally through the low mobility molecular film, reducing the likelihood of discharge through the oxide defect.

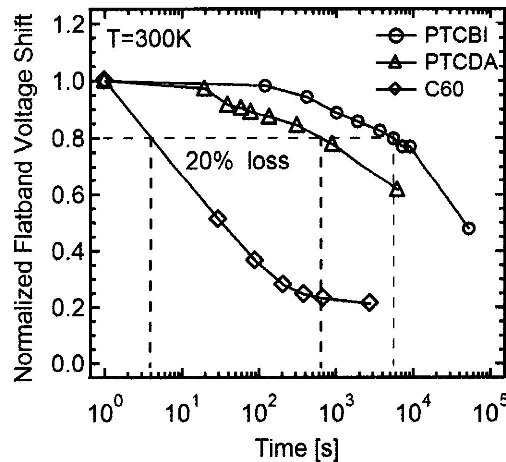


Figure 2-13: Normalized retention characteristics measured at room temperature [44].

TABLE II

20% CHARGE LOSS TIME VERSUS ELECTRON LATERAL MOBILITY

	PTCBI	PTCDA	C_{60}
Electron Mobility (cm^2/Vs)	2.4×10^{-6}	10^{-5}	5.1×10^{-2}
20% Charge Loss Time (s)	5.4×10^{3}	6.6×10^{2}	4

From the charge retention results on the set of the tested molecular thin film memories we composed Table II, which shows the dependency of the retention time on the lateral electron mobility of the molecular layer in these memory devices. A clear trend of the reduced memory retention time with increased thin film mobility is apparent, confirming that bulk of charge retention is in the molecular floating gate film rather than in the trap states of the structure.

Charge retention of these devices can be further improved by utilizing a higher quality dielectric film, by replacing the sputtered Al_2O_3 layers with high quality Al_2O_3 layer deposited by atomic-layer deposition (ALD) system or by applying a bi-layer tunnel dielectric consisting of high dielectric constant oxides such as HfO_2.

Conclusion

In conclusion, we investigated the memory properties of a set of organic materials for possible application to CMOS-compatible nonvolatile memory devices. High charge retention density and good cycling endurance was demonstrated in PTCDA, PTCBI and C_{60} structures. Also, charge retention dependency on the lateral mobility of organic thin films has been demonstrated, suggesting use of low mobility materials like PTCBI which consists of inherently well-ordered planar molecules, with lateral mobility as low as $10^{-6}cm^2V^{-1}s^{-1}$, high molecular area density and thermal stability (>350°C) which makes it a viable option for the floating gate of flash memory devices.

These results may lead to an approach toward further miniaturization of non-volatile memory by using molecules as segmented charge storage elements in the floating gate flash memory technology.

Bibliography

[1] P. Pavan, R. Bez, P. Olivo, E. Zanoni, "Flash Memory Cells-An Overview," Proc. of the IEEE., vol. 85, no. 8,pp. 1248-1271, 1997.

[2] M. She, "Semiconductor Flash Memory Scaling," Ph.D. thesis, University of California, Berkeley, fall 2003.

[3] Seiichi Aritome, "Advanced Flash Memory Technology and Trends for Files Storage Application", pp.763, *IEDM 2002*.

[4] D.J. Jung, "Highly Manufacturable 1T1C 4Mb FRAM with Novel Sensing Scheme", pp.279-282, *IEDM* 1999

[5] "Advanced Memory Technology and Architecture", short course, *IEDM* 2001.

[6] S. Lai and T. Lowrey, "OUM- A 180nm Nonvolatile Memory Cell Element Technology for Stand Alone and Embedded Applications", pp.803, *IEDM 2001*.

[7] Paolo Cappelletti, "Flash Memories", *Kluwer Academic Publishers*, 1999.

[8] International Technology Roadmap for Semiconductors, ITRS, 2007. Available: http://www.itrs.net

[9] E. Suzuki, H. Hirashi, K. Ishii, and Y. Hayashi, "A low-voltage alterable EEPROM with metal-oxide-nitride-oxidesemiconductor (MONOS) structure," IEEE Trans. Electron Devices, vol. ED-30, no. 2, pp. 122–127, 1983.

[10] E. Suzuki, K. Miura, Y. Hayashi, R.-P. Tsay, and D. Schroder, "Hole and electron current transport in metal-oxide-nitrideoxide-silicon memory structures," IEEE Trans. Electron Devices, vol. 36, no. 6, pp. 1145–1149, 1989.

[11] P. E. Cottrell, R. R. Troutman, and T. H. Ning, "Hot-electron emission in n-channel IGFET's," IEEE Trans. Electron Devices, vol. ED-26, no. 4, pp. 520–532, 1979.

[12] B. Eitan and D. Froham-Bentchkowsky, "Hot-electron injection into the oxide in n-channel MOS devices," IEEE Trans. Electron Devices, vol. ED-28, no. 3, pp. 328–340, 1981.

[13] H. Abdu, "Molecular and quantum dot floating gate non-volatile memories," Master of Engineering Thesis, Massachusetts Institute of Technology (June 2008);

[14] R. Bez, E. Camerlenghi, A. Modelli, A. Visconti, "Introduction to Flash Memory," proceedings of the IEEE, VOL. 91, NO. 4, April 2003.

[15] P. Cappelletti, R. Bez, D. Cantarelli, and L. Fratin, "Failure mechanisms of Flash cell in program/erase cycling," in IEDM Tech. Dig.,1994, pp. 291–294.

[16] F. Masuoka, M. Momodomi, Y. Iwata, and R. Shirota, ''New Ultra High Density EPROM and Flash EEPROM with NAND Structure Cell,'' IEDM Technical Digest, 1987, pp. 552–555.

[17] S. K. Lai, "Flash memories: successes and challenges," IBM J. Res. & Dev., vol. 52, no. 4-5, pp. 529-535, 2008.

[18] D. Kwak, J. Park, K. Kim, Y. Yim, S. Ahn, Y. Park, J. Kim, et al., ''Integration Technology of 30nm Generation Multi-Level NAND Flash for 64Gb NAND Flash Memory,'' Symposium on VLSI Technology, Digest of Technical Papers, Kyoto, Japan, 2007, pp. 12–13.

[19] M. Wei, R. Banerjee, L. Zhang, A. Masad, S. Reidy, J. Ahn, H. Chao, et al., ''A Scalable Self-Aligned Contact NOR Flash Technology,'' Symposium on VLSI Technology, Digest of Technical Papers, Kyoto, Japan, 2007, pp. 226–227.

[20] G. Atwood, ''Future Directions and Challenges for ETox Flash Memory Scaling,'' IEEE Trans. Device Material Rel. 4, No. 3, 301–305 (2004).

[21] H. B. Pein and J. D. Plummer, ''Performance of the 3-D Sidewall Flash EPROM Cell,'' Proceedings of the International Electron Devices Meeting, IEDM Technical Digest, December 5–8, 1993, pp. 11–14.

[22] K. Prall, ''Scaling Non-Volatile Memory below 30nm,'' NonVolatile Semiconductor Memory Workshop, August 26–30, 2007, pp. 5–10.

[23] J.Jung, W. J. Cho"Tunnel Barrier Engineering for Non-Volatile Memory," journal of semiconductor technology and science, VOL.8, NO.1, MARCH, 2008

[24] D. Wang, T.-P. Ma, J. W. Golz, B. L. Halpern, and J. J. Schmitt, ''High-Quality MNS apacitors Prepared by Jet Vapor Deposition at Room Temperature,'' IEEE Electron Device Lett. 13, No. 9, 482–484 (1992).

[25] Y. Shin, J. Choi, C. Kang, C. Lee, K.-T. Park, J.-S. Lee, J. Sel, et al., ''A Novel NAND-Type MONOS Memory Using 63 nm Process Technology for Multi-Gigabit Flash

EEPROMs,'' IEEE Electron Devices Meeting, IEDM Technical Digest, December 5–7, 2005, pp. 327–330.

[26] A. Fazio and M. Bauer, ''Intel StrataFlashe Memory Technology Development and Implementation,'' Intel Technol. J. 1, No. 2, Q4 (1997).

[27] International Technology Roadmap for Semiconductors, ITRS, 2007. Available: http://www.itrs.net

[28] K. Kinam, "Technology for sub-50nm DRAM and NAND Flash Manufactureing," IEDM, 1609340, pp. 323-326, 2005.

[29] P. K. Singh, R. Hofmann, K. K. Singh, N. Krishna, and S. Mahapatra "Performance and reliability of Au and Pt single-layer metal nanocrystal flash memory under NAND (FN/FN) operation," IEEE Trans. Electron Devices, vol. 56, no. 9, Sep. 2009.

[30] O. Nayfeh, D. Antoniadis, K. Mantey, M. Nayfeh, "Memory effects in metal-oxide-semiconductor capacitors incorporating dispensed highly monodisperse 1nm silicon nanoparticles," Appl. Phys. Lett., vol. 90, pp. 153105, 2007.

[31] J.-S. Lee, Y.-M. Kim, J.-H. Kwon, H. Shin, B.-H. Sohn, and J. Lee, "Tunable memory characteristics of nanostructured, nonvolatile charge trap memory devices based on a binary mixture of metal nanoparticles as a charge trapping layer," Adv. Mater., vol. 21, pp. 178–183, 2009.

[32] S. Forrest, "Ultrathin organic films grown by organic molecular beam deposition and related techniques," Chem. Rev, vol. 97, pp. 1793-1896, 1997.

[33] S. H. Cho, D. I. Lee, J. H. Jung and T. W. Kim, "Electrical bistabilities and memory stabilities of nonvolatile bistable devices fabricated utilizing C_{60} molecules embedded in a Polymethyl Methacrylate layer," Nanotechnology, vol. 20, pp. 345204, 2009.

[34] S. H. Kang, T. Crisp, I. Kymissis, and V. Bulovic, "Memory effect from charge trapping in layered organic structures," Appl. Phys. Lett., vol. 85, no. 20, 15 November 2004.

[35] J. C. Scott and L. D. Bozano, "Nonvolatile memory elements based on organic materials," Adv. Mater., vol. 19, pp. 1452–1463, 2007.

[36] E. I. Haskal, Z. Shen, Z., P. E. Burrows, S. R. Forrest, "Excitons and exciton confinement in crystalline organic thin films grown by organic molecular-beam deposition," Phys. Rev. B, vol. 51, pp. 4449, 1995.

[37] B. P. Rand, J. Genoe, P. Heremans, J. Poortmans, "Solar cells utilizing small molecular weight organic semiconductors," Prog. Photovolt: Res. Appl. vol. 15, pp.659–676, 2007.

[38] J. R. Ostrick, A. Dodabalapur, L. Torsi, A. J. Lovinger, E. W. Kwock, T. M. Miller, M. Galvin, M. Berggren, H. E. Katz, "Conductivity-type anisotropy in molecular solids," J. Appl. Phys., vol. 81, pp. 6804, 1997.

[39] G. G. Malliaras, Y. Shen, and D. H. Dunlap, "Nondispersive Electron Transport in Alq3," Appl. Phys. Lett., vol. 79, no. 16, pp. 2582, 2001.

[40] C. Madigan, V. Bulovic, "Exciton energy disorder in polar amorphous organic thin films: Monte Carlo calculations," Phys. Review B, vol. 75, pp. 81403, 2007.

[41] C. Madigan, V. Bulovic, "Exciton energy disorder in polar amorphous organic thin films: Monte Carlo calculations," Phys. Review B, vol. 75, pp. 81403, 2007.

[42] S. R. Forrest, M. L. Kaplan, P. H. Schmidt, "Organic-on-inorganic semiconductor contact barrier diodes. I. Theory with applications to organic thin films and prototype devices," J. Appl. Phys., vol. 55, pp. 1492, 1984.

[43] Kendrick, C.; Kahn, A.; Forrest, S. R. Appl. Surf. Sci. 1996, 104/105, 586

[44] S.Paydavosi, H.Abdu, G. J. Supran, V.Bulovic, "Performance Comparison of Different Organic Molecular Floating Gate Memories,"IEEE Transactions on Nanotechnology, 2011 (accepted)