Proceedings

Non-Volatile Memory Technology Symposium 2000

http://nvm.jpl.nasa.gov

November 15-16, 2000

Executive Center, Strategic Analysis, Inc.
One Virginia Square, Arlington, Virginia

Sponsors:

JPL

CISM

Proceedings

Non-Volatile Memory Technology Symposium 2000

http://nvm.jpl.nasa.gov

November 15-16, 2000

Executive Center, Strategic Analysis, Inc.
One Virginia Square, Arlington, Virginia

Hosted by
  Center for Integrated Space Microsystems (JPL)
  Remote Exploration and Experimentation Project (JPL)

Sponsored by
  National Aeronautics and Space Administration (NASA)
  Defense Advanced Research Projects Agency (DARPA)
  Defense Threat Reduction Agency (DTRA)
  Air Force Research Laboratory (AFRL)

Sponsors:

JPL

Non-Volatile Memory Technology Symposium 2000

November 15-16, 2000

Executive Center, Strategic Analysis, Inc.
One Virginia Square, Arlington, Virginia

NVMTS 2000 Organizing Committee

General Chair
Taher Daud (JPL)

Co-Chairs
Lewis Cohn (DTRA)
Stuart Wolf (DARPA)

Program and Technical Chair
Karl Strauss (JPL)

Program and Technical Committee
Leon Alkalai (JPL)
Nazeeh Aranki (JPL)
Charles Barnes (JPL)
Gary Derbenwick (Celis Semiconductor)
Kenneth Hunt (AFRL)
Sammy Kayali (JPL)
Jagdish Patel (JPL)
Raphael Some (JPL)
Scott Tyson (AFRL)

Welcome Message

Welcome to the Nonvolatile Memory Technology Symposium 2000. This first symposium provides an open forum for the presentation of myriad advances in the nonvolatile memory technology during the recent past. It covers not only the cell design and the applications, but also the related issues of radiation environment and die packaging.

The Symposium was put together in a very short time which kept the organizers on their toes, and kept the pressure of quick turnover from the contributors. A well-deserved credit of all the hard work goes to the organizers and the contributors.

The Symposium was founded with the basis of being open to all NVM technologies, not just focusing on one or the other – cross-pollination will occur.

As with any fledgling enterprise, certain bumps will occur on the road to success. We thank our Sponsors and Supporters for their substantial assistance. With DARPA's help, we were able to obtain this Executive Center. Sized for about 100 people, we consider to be just right for us. It is located close to Metro and affords easy access with the flexibility for attendees for lodging around town to their convenience.

A total of 28 talks, both contributed and invited, have made this program for 1 1/2 days a great success. It has, we hope, the right breadth and depth to keep you all keenly interested and absorbed. During the three half periods, three invited talks, each covering a different area are being presented. We hope these talks will prove to be specially useful for users and innovators alike. We have planned four sessions. Session I has papers related with memory cell design and simulations. Session II on design challenges and applications, has the largest contributions, with 9 papers, which include papers on radiation environment. Sessions III is devoted to innovative concepts with 7 papers. On the following day, Session IV deals with newer, emerging memory technologies.

Many thanks to all the contributors without whose diligence and active participation no conference can be successful. Also acknowledged is the hard work of the committee members and organizers for taking special efforts at such a short notice.

We hope you will find the symposium stimulating as well as rewarding and helpful. We surely would appreciate receiving your feedback in terms of comments, suggestions, criticisms on all aspects of the conference. Please feel free to formally or informally let us know of your preferences and things to be included, if any, in our future symposiums.

Taher Daud   Karl Strauss
General Chair   Program and Technical Chair

TIME	Wednesday, 15 November 2000
8 30 AM	Welcome
8 40 AM – 9 00 AM	Invited - Design Challenges of the Europa Orbiter Mission; Robert Staehle, Jet Propulsion Laboratory

Session I - Memory Cell Design & Simulation

TIME	
9 20 AM	Implementation of Self-Align Local Poly Interconnection on 0.25 um Flash Memory; CH Lee, et al, United Microelectronics
9 40 AM	A 0.25um Embedded Flash Technology with Shallow Trench Isolation Using Channel FN Operation; CJ Huang, et al, United Microelectronics
10 00 AM	Design Considerations in Scaled SONOS Nonvolatile Memory Devices; J Bu, et al, Lehigh University
10 20 AM	(break)
10 40 AM	3-D Computer Simulation of Nano-crystal Floating Gate Flash Memory Devices; A Thean, University of Illinois, Champaign

Session II - Challenges & Applications

TIME	
11 00 AM	Challenges in Implementing Commercial Non-Volatile Memory in Spacecraft Solid State Recorders; L Adams, et al, SEAKR Engineering
11 20 AM	Expectations of MRAM in comparison with other non-volatile memory technologies; K-M Lenssen, Philips Research Laboratories
11 40 AM	Opportunities and Challenges for Embedded Flash Memory; S Kianian, et al, Silicon Storage Technology Inc.
12 00 PM – 12 40 PM	LUNCH
1 00 PM – 1 20 PM	Invited - Radiation Factors of Reliability on Non-Volatile Electronics; Sammy Kayali, Jet Propulsion Laboratory
1 40 PM	Radiation Issues and Applications of Floating Gate Memories; L Scheick, Jet Propulsion Laboratory
2 00 PM	Total Dose Radiation Response and High Temperature Imprint Characteristics of Chalcogenide Based RAM Resistor Elements; S Bernacki, Raytheon Systems
2 20 PM	SONOS Non-Volatile Semiconductor Memories for Space and Military Applications; D Adams, Northrop Grumman
2 40 PM	SONOS Non-Volatile Shadow RAMs for Space Applications; G Derbenwick, Celis Semiconductor
3 00 PM	Implementation of FeRAM for Space Applications; S Philpy, Celis Semiconductor
3 20 PM	(break)

Session III - Innovative Concepts

TIME	
3 40 PM	Characterization of a New EEPROM/Flash Memory Cell; R Reedy, Peregrine Semiconductor
4 00 PM	SOI Non-volatile Memory; P McMarr, Naval Research Laboratory
4 20 PM	Mechanisms of Protonic Nonvolatile Memory Device; P Macfarlane, Naval Research Laboratory
4 40 PM	Functional Nanostructured Particle Arrays for Nanocrystal Nonvolatile Memories; H Atwater, California Institute of Technology
5 00 PM	Quantitative Analysis of Charge Injection and Discharging of Si Nanocrystals and Arrays by Atomic Force Microscopy; LD Bell, Jet Propulsion Laboratory
5 20 PM	Advanced Compact Holographic Data Storage System; T-H Chao, Jet Propulsion Laboratory
5 40 PM	Ultra-High Density Holographic Memory Module in Solid-State Architecture; V Markov, Metrolaser
6 00 PM	End of Day Remarks

TIME	Thursday, 16 November 2000
8 30 AM	Welcome
8 40 AM – 9 00 AM	Invited - Sources of Government Funding for New Developments; Ken Hunt, Air Force Research Laboratory

Session IV - New Technologies

TIME	
9 20 AM	Chalcogenide-Based Non-Volatile Memory; J Maimon, Ovonyx
9 40 AM	Advances in FeRAM Technology; G Derbenwick, Celis Semiconductor
10 00 AM	High Density Packaging of Non-Volatile Memory Systems; J Carson, Irvine Sensors
10 20 AM	(break)
10 40 AM – 11 00 AM	All Metal Magnetic RAM; E Torok, IME
11 20 AM – 11 40 AM	Advanced MRAM Concepts; J Daughton, Nonvolatile Electronics
12 00 PM – 12 40 PM	CLOSING REMARKS and LUNCH
1 00 PM	End of Symposium

NVMTS2000 Schedule / Program

v

Executive Center Conference Facilities
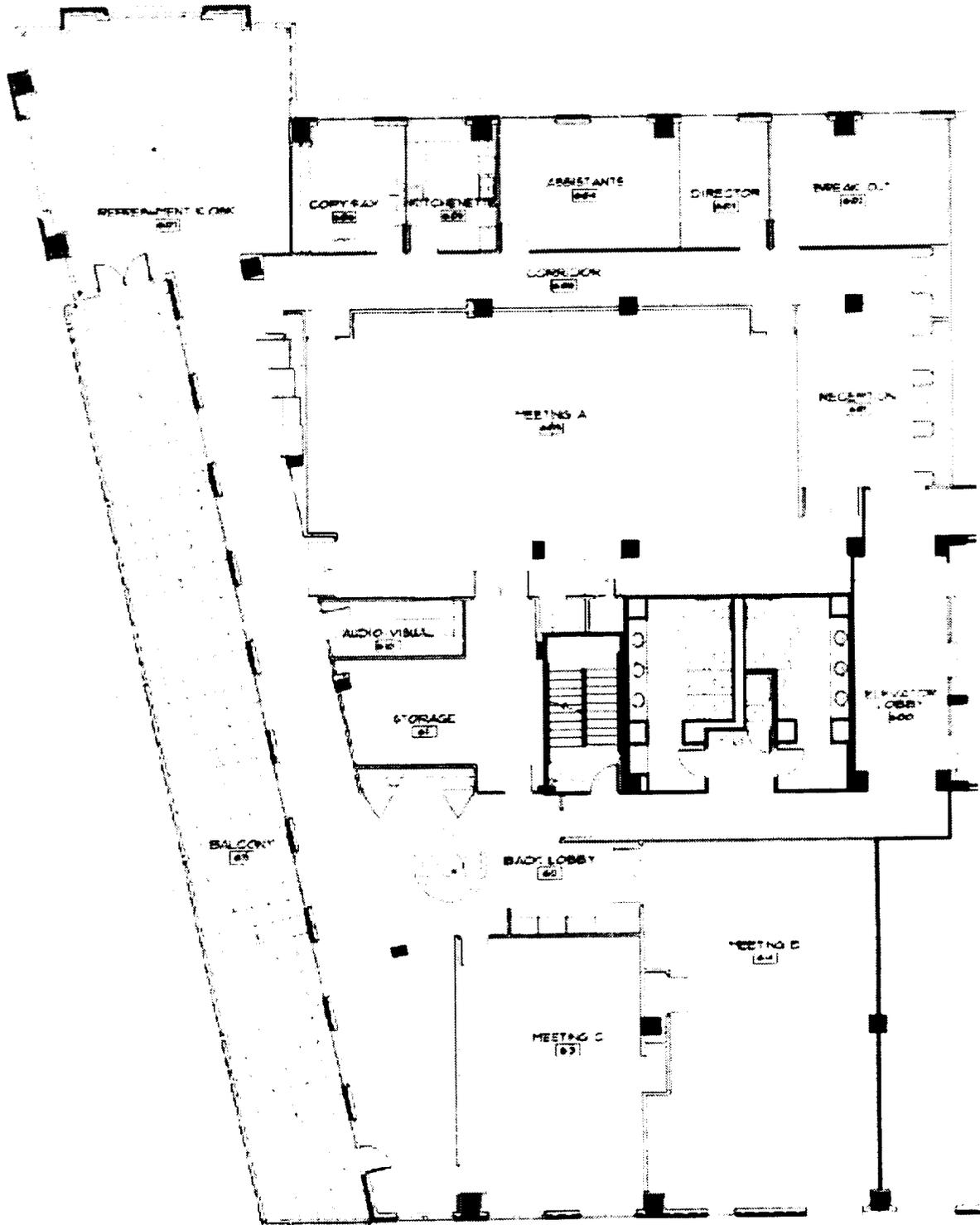
Map of Surrounding Area
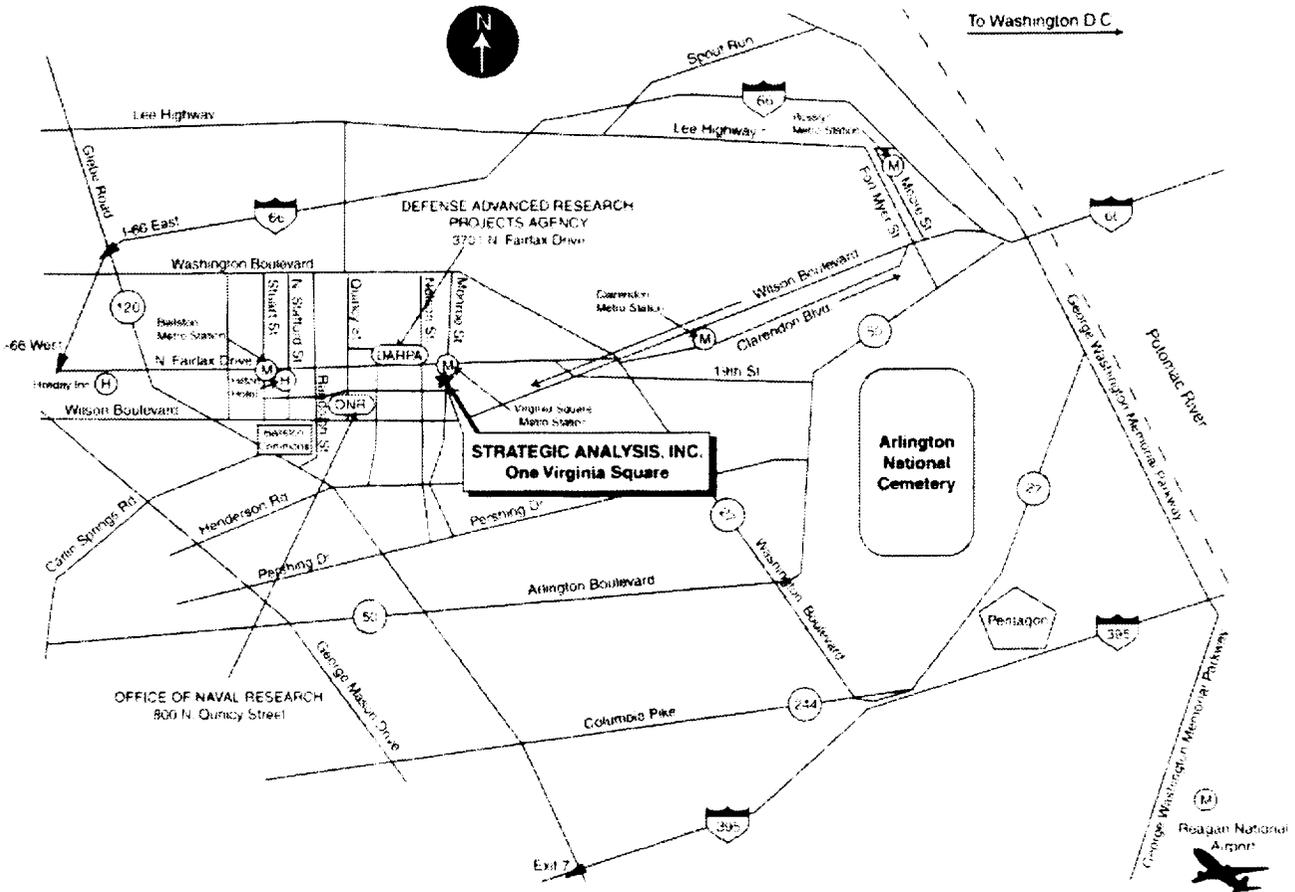
Invited Talks

Wednesday, November 15

    8:40 AM   Design Challenges of the Europa Orbiter Mission;
                        Robert Staehle, Jet Propulsion Laboratory

    1:00 PM   Radiation Factors of Reliability on Non-Volatile Electronics;
                        Sammy Kayali, Jet Propulsion Laboratory

Thursday, November 16

    8:40 AM   Sources of Government Funding for New Developments;
                        Ken Hunt, Air Force Research Laboratory

SESSION #1: *Memory Cell Design and Simulation*

Wednesday, November 15, 2000
9:20 AM–11:00 AM
Chair: *Gary Derbenwick*, Celis Semiconductor,
Colorado Springs, Colorado

9:20 AM   Implementation of Self-Align Local Poly Interconnection on 0.25 um Flash Memory; *C.H. Lee, C.H. Lin, H.H. Chen, Y.K. Sheu, W.C. Ting and G. Hong*, United Microelectronic, Taiwan

9:40 AM   A 0.25um Embedded Flash Technology with Shallow Trench Isolation Using Channel FN Operation; *C.J. Huang, Y.C. Liu, P.C. Lin, A. Wu, H.H. Chen, W.C. Ting and G. Hong*; United Microelectronics, Taiwan

10:00 AM   Design Considerations in Scaled SONOS Nonvolatile Memory Devices; *J. Bu and M. White*, Lehigh University

10:20 AM   (Break)

10:40 AM   3-D Computer Simulation of Nano-crystal Floating Gate Flash Memory Devices; *A. Thean and J. Leburton*, University of Illinois - Urbana, *M. Sadd*, Motorola

Implementation of Self-Align Local Poly Interconnection on 0.25um Flash Memory

C.H. Lee. C.H. Lin, H.H. Chen, Y.K. Sheu, WenChi Ting, Gary Hong
Specialty Technology Division.
United Microelectronic Corp.
No. 3, Li-Hsin Rd. 2, Science-Based Industrial Park, Hsinchu, Taiwan, ROC

Abstract-For ETOX Flash technology, the self-aligned source module plays a critical role in minimizing cell size. For advanced Flash technology using the shallow-trench isolation (STI) scheme, however, it is difficult to form a stable source line. The Flash process in this experiment features an ETOX structure and with SIN spacer. Poly-Si (implanted poly) is used as the local interconnection material connecting the source line and was successfully implemented in the ETOX flash array in this paper.
Keyword: ETOX, local interconnection, shallow trench isolation (STI)

TABLE OF CONTENTS

1.INTRODUCTION:

Common ground line, like word line and bit line is the base component of the flash memory structure (NAND, NOR, DINOR...). Since the ground node is necessary when operating a flash cell (ex: program, erase, or read). To minimize the array size, the ground node should be connected as a common ground line, especially for high-density memory application. In the 0.35um technology, the ground node side's LOCOS was etched away and formed a stable implanted line after proper implants. This is because of being with a rather smooth active corner under the isolation comparing with the STI isolation. For 0.25um technology, LOCOS couldn't meet the more tough design rule and thus the STI process became the best choice. Nevertheless, it is difficult to form a stable source line using the STI process. Many methods are proposed to solve this problem, such as changing the tilt angle of implant, fine tuning the STI structure to a bow like structure to solve the unstable issue or just put a common active line to connect each ground node which however would suffer the cell size. In this paper, we use the poly as the common ground line. It's the first try to use another material to connect every junction instead of implants.

2.EXPERIMENTS

The flash in the experiment is an ETOX structure with 64*64bits array. We extracted a single cell in this array and check it's characteristic. The cell size is 0.9775um^2 based on cell channel length of 0.5um and channel width of 0.32um. TunOx is furnace thermal oxide with thickness 100Å. The effective thickness of ONO inter-poly dielectric is 150Å. Fig.1 show the X-view SEM photograph of the final flash cell. The source side is connected to the local poly interconnection. The drain side is connected to the bit line and pick-up by the tungsten plug contact. To sustain higher junction voltage and increase the program speed, the drain side junction is implant by phosphorous 20kev 3e15cm^2 implants in addition to N+ and LDD implants. The local poly interconnection is phosphorous implant poly with dosage of 4e15cm^2 and the sheet resistance is 99_/sq

The Flash process in this experiment features an ETOX structure and with SIN spacer. Poly-Si is used as the local interconnection material connecting the source line. First of all, the common source line was etched using self-align etching, so we could define a fine trench line. Secondly, the poly refilled the etched source line. Finally, the poly interconnect is defined after photo and etching.

3.RESULTS

The final X-view SEM photograph is shown as Fig.1. We would like to list the highlights in this paper.

A SiN/SiON stack is deposited on the P2. Then the SiN/SiON/P2/ONO/P1 sandwich structure is defined in the following p2 etching step. The cap SIN in this experiment is 1.5k Å and SiON is 300Å. After that, we use SiN spacer (1000 Å) to cap the P2 and P1.

The selectivity of oxide to P2 cap SiN and SiN spacer is difficult to control, since the thin top pad TEOS oxide between SiN spacer and cap SiN is a line structure whose area is very large. The expose pad oxide provides too much oxygen during etch to reduce polymer and thus degrade the oxide to SiN selectivity. Fig.2 shows the P3 short to P2 by using tuned dry etching recipe. We use the oxide wet etching back process after IPD deposition and proper dry etching after self-align source photo to solve this problem. Please see the tilt-SEM photograph as shown in the Fig.3

The topography after self-align source etching is very rough and is thus prone to poly residue problems. We had overcome the residue issue by fine tuning the over-etching recipe. The resulted photograph is shown as in the Fig.4

The basic Program/Erase characteristic is shown as in the Fig.5. The erase condition of Vg is 19V and the other nodes are ground. The program conditions are Vd=5V,Vg= -12V, Vs= Open and substrate is ground

4.CONCLUSION:

We had successfully implemented the local poly interconnection in the one transistor flash cell. We had overcome some process integration difficulty such as self-align align source etching and Poly3 residual issue. It's the first try and a good start to use poly as the connection material. We could also use tungsten or Ti/TiN as the connection material in the future.

5.ACKNOWLEDGMENT:

The authors would like to thank Kraus Wu for the layout support and also thank Ashley Tsai, Sinead Wu and Kao Su Huang for their kind efforts for solving related etching problem. Finally, I would thanks my wife for the encouragement throughout this work.

6.REFERENCE:

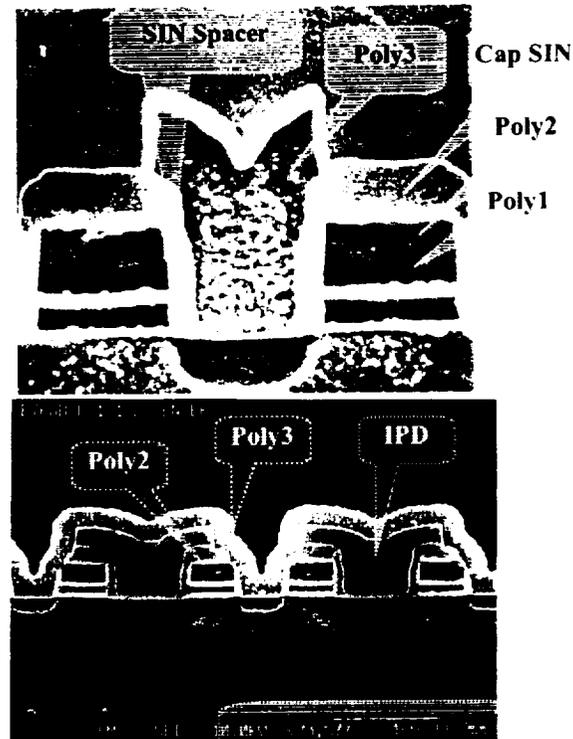[1].Willian D.Brown,. "Nonvolatile Semiconductor Memory Technology"
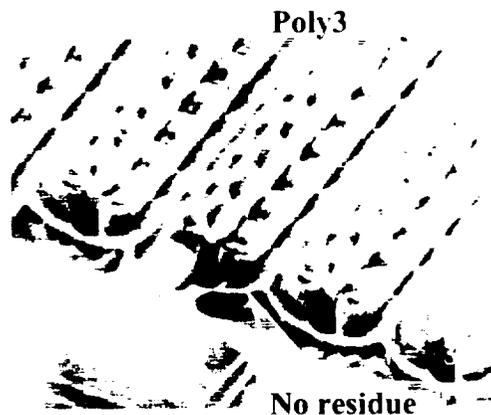
Fig.3 Self align source after proper etch

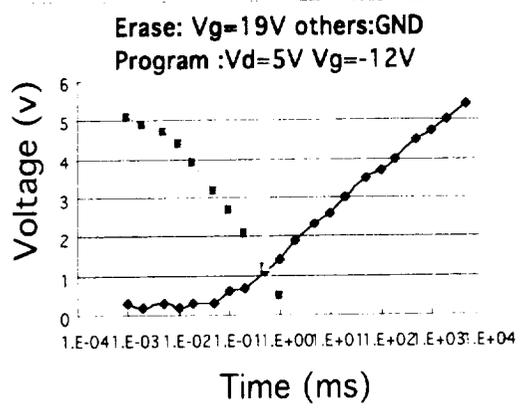Fig.4 The resulted tilt-photograph after poly3 etching.

Fig.5 The basic Program/Erase characteristic

A 0.25μm Embedded Flash Technology with Shallow Trench Isolation Using Channel FN Operation

C.J. Huang, Y.C. Liu, P.C. Lin, A Wu, H.H. Chen, W.C. Ting and Gray Hong

Specialty Technology Division, United Microelectronics Corporation,
No. 3, Li-Hsin Rd. 2, Science-Based Industrial Park, Hsin Chu, Taiwan, R.O.C.
E-mail: c_j_huang@umc.com, TEL:886-3-5782258 ext 37031, FAX:886-3-5797246

Abstract— A new 0.25 μm flash EEPROM developed for embedded applications will be reported. Flash memory is achieved by utilizing a single transistor NOR type cell that employs Fowler-Nordheim tunneling for both program and erase operations. Channel FN tunneling is a high efficiency and low power consumption approach for flash cell operation. The flash EEPROM is integrated into a standard quarter micron logic process with dual gate oxide for high performance and high voltage transistors. Program and erase could be achieved within 5 ms to target. The high performance logic devices didn't be degraded by incorporated with flash process was demonstrated.

TABLE OF CONTENTS

1. INTRODUCTION

Combining high density, low cost and electrically programmable and erasable capabilities, flash EEPROM has been widely accepted as the memory of choice for program and data storage. Embedded flash EEPROM can offer enhanced system performance and added many advantages [1]. A robust flash EEPROM cell compatible with the existing host logic process without need for tight process controls is a key criterion for use in embedded applications. Conventional NOR-type flash memory use channel hot electron for program which has large power consumption and FN tunneling with reversed biased deep junction for erase which impact the cell shrink. In this paper, we describe the feasibility of the flash EEPROM cell using FN tunneling mechanism for programming and erasure operations. Tight threshold voltage distribution can be achieved by using bit-by-bit verification during programming. It also provides a solution for disturb free operation and low power consumption.

In this paper, we demonstrate the feasibility of the flash EEPROM cell, which uses FN tunneling mechanism for both programming and erasure operations without the sacrifice of the high performance of the logic devices. It also provides a solution for low disturb and low power consumption.

2. FLASH CELL STRUCTURE and OPERATION

The flash memory employs triple well, double layers of polysilicon and three layers of metal based on 0.25μm CMOS process. The flash cell developed in this technology is a single transistor NOR type cell with separate drain and source line for each column. Fig. 1 shows the cross-section view with TEM picture of memory cell along the bit-line and word-line direction.

(a)   (b)

Fig.1 The cross-sectional TEM view of the flash cell along (a) the word-line and (b) bit-line direction.

The effective thickness of tunnel oxide and ONO is 90C and 175C, respectively. The key process parameters are described in Table I.

Table I. Key parameter of Cell

Items	Size
Cell gate length	0.34 um
Cell Channel width	0.25 um
Floating gate on STI	0.19 um
Contact to Poly	0.19 um
Trench depth	0.4 um
Tunnel oxide	9.0 nm
Interpoly dielectric (ONO)	17.5 nm

An experimental bit cell with channel length of 0.34 μm and channel width of 0.25 μm is constructed for characterization. The UV Vt of the flash cell is ~ 2.2V

and GCR of ~ 0.56. The process flow is based on standard 0.25 μm logic ones to do some modification. We add DNW/HVPW process for cell and HV devices. The tunnel oxide growth, floating gate poly deposition and patterning and ONO formation for flash cell were process after STI formation. All high temperature thermal budgets are completed before formation of the logic devices in order not to impact the logic devices' performances. We also need a thick gate oxide for high voltage devices to implement the flash cell operations. We define the program to mean putting electrons onto floating gate and erase to mean removing electrons from the floating gate [2]. Therefore, the cell had high cell V_T after program and low V_T after erase. Table II summarized the cell operation bias conditions of this flash cell technology.

Table II. The cell operation Bias conditions table

		Drain	Gate	Source	Bulk
Program	Select	-6.0	+10.5	F	-6.0
	Unselected	GND	GND	F	-6.0
Erase	Select	F	-11.5	F	+7.0
	Unselected	GND	GND	F	+7.0
Read	Select	1.0	3.0	GND	GND

The equivalent oxide thickness was measured by high frequency C-V with HP 4280A. The tunnel oxide physical thickness was confirmed by TEM pictures. The flash cell performances were measured by HP4156B, HP8110A and software developed by ourselves. The peripheral devices were checked by HP4071 system.

3. RESULTS and DISCUSSION

The tunnel oxide quality with long and short STI edges were estimated by charge to breakdown (Q_{BD}) test
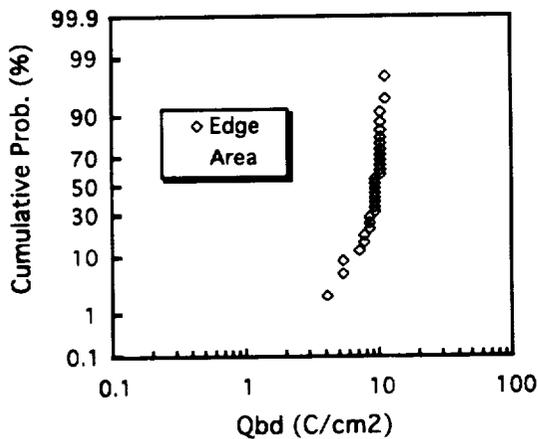

Fig.2. The tunnel oxide Q_{BD} characteristics for the different STI edged pattern

and shown in fig.2. The Q_{BD} distribution of long STI edge didn't be degraded comparing to that of area one. It shows our STI module which uses special corner rounding process didn't impact the tunnel oxide quality.

Programming operation of flash cell occurs by channel FN tunneling from the floating gate to the inverted channel, which results in high threshold voltage state. It performed by applying +10.5 V on the control gate (CG), -6.0 V on the drain and N-well, and source floating. The cell is programmed using channel FN tunneling and a typical characteristic is shown in fig.2. For these bias condition, the target of $V_T > +4.5$ V could be achieved within 5 ms.
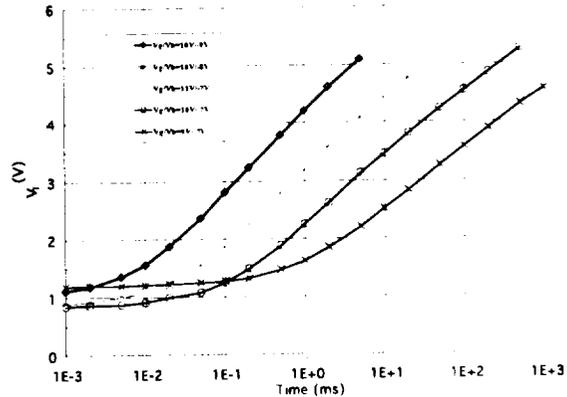
Fig.3 The programming characteristics of flash cell with STI edge via channel FN tunneling

The erase operation of flash cell by channel FN tunneling pull the electrons out of FG with bias conditions of -11.5 V on CG, +7.0 V on N-well, drain and source floating, respectively. We illustrated erase characteristics of the cell in fig.3. For the given bias condition, the target of $V_T < +1.5$ V is attained in < 5 ms.
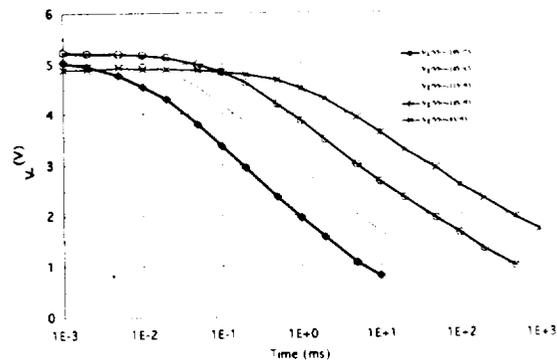
Fig.4 The erase characteristics of flash cell with STI edge via channel FN tunneling

Because the erase operation involves tunneling to a channel at a uniform potential, a high density of hot holes are not created at a reverse biased junction which avoids the reliability issues that have been reported on approaches that have strong band-to-band drain tunneling. [3,4] Because only FN tunneling is used during programming and erasure operations, low power consumption is achieved. The steep subthreshold slope shows that this cell had good off-current performance.

6

The cell current characteristics under read operation condition with +3.0 V on the CG, and +1.0 V on the drain are measured. The Ids is ~ 25 µA at the nominal cell read bias and it is higher than our specification (> 20 µA). The program inhibit capability were checked. The data was shown in fig.5. We stressed the erased cells with different CG voltage with bulk fixed at -6V and drain grounded. The program inhibit capability is excellent for this flash cell even up to 10 s disturbance time. The optimization of the drain junction needs to be taken care for this program inhibit operation.

Fig 5 The program inhibit characteristics of the channel FN P/E flash cell. Negligible VT shift is observed up to 10 s

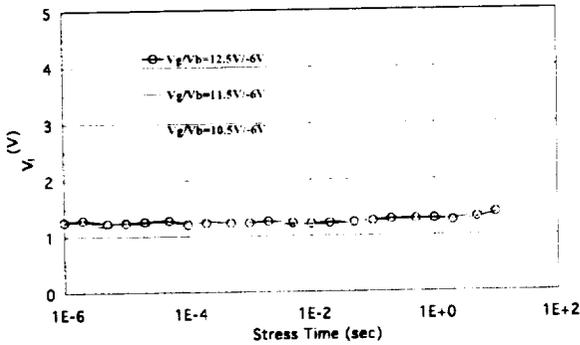Susceptibility to disturb during erasure was checked by stressing the programmed cells with 6, 7 and 8 V on the bulk, 0 V on the CG and drain/source floating. No disturb was observed even up to 10 sec stress time for the selected word line cells as shown in fig.6.
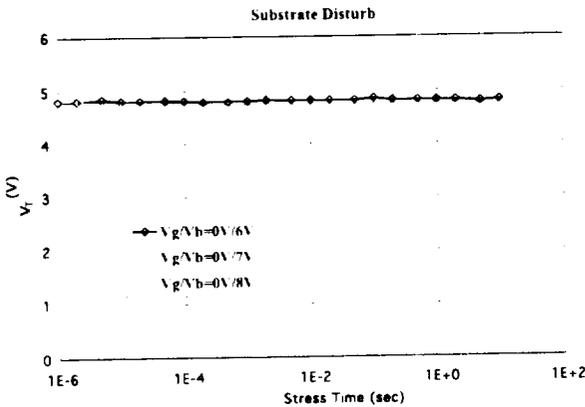
Fig 6 The substrate disturb characteristics during erase operation. Negligible Vt shift is observed up to 10 sec

A program/erase endurance of 10k cycles is demonstrated for a single cell at room temperature as shown in fig.7. The operation bias conditions are same as the one stated previously. After 10k endurance, we didn't observe the V_T window closure but instead window shifting upward. After cycling, we found that the program speed is faster and erase speed is degraded compared to a virgin cell. The electron traps near the FG and hole traps near the

substrate were the root cause to explain this abnormal behavior [5,6].
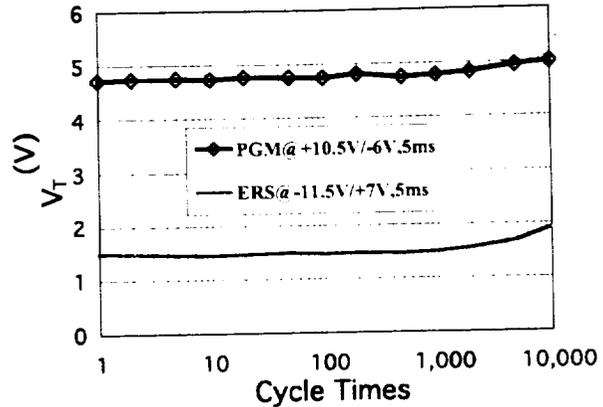
Fig.7. The endurance characteristics of the flash single cell at room temperature use channel FN tunneling for both program and erase operation.

The data retention tests of programmed and erased single cells with baking temperature of 250C were performed and shown in fig.8. The cell V_T was measured at baking time of 24hr and 168hr. Negligible charge loss or charge gain was observed. This demonstrates the inherent capability of this cell to retain charges.
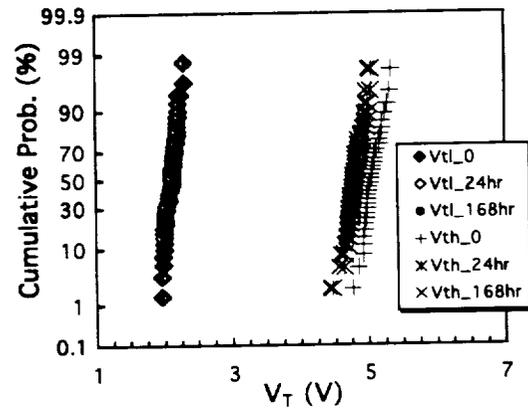
Fig.8. The data retention data of the flash cell baked at 250C withstanding 24 and 168 hrs. Negligible charge loss or gain is observed.

Fig.9 (a) and (b) illustrate the I_{on}/I_{off} characteristics of the 50Å low voltage NMOS and PMOS transistors.

LV-NMOS Ids vs Ioff

(a)

respectively.

LV-PMOS Ids vs Ioff
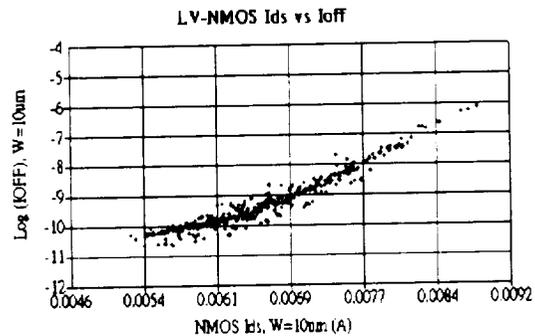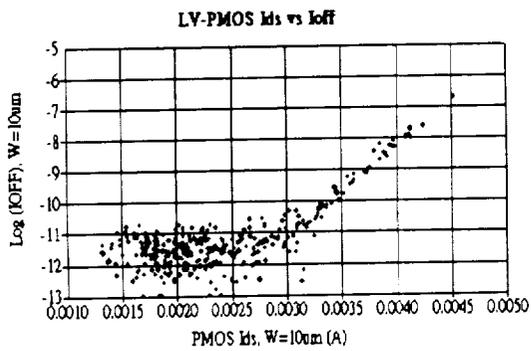
(b)

Fig.9. (a). The LVNMOS universal curves (b) The LV-PMOS
universal curves. These show that the Ids meet the spec
(N:600uA/um, P:280uA/um) at the requirement (Ioff=1nA/um)

Compared to a pure logic process with no embedded flash
modules using same gate oxide thickness, no degradation
in transistor performance is observed.

4. Conclusion

In summary, we had demonstrated a new cell
technology with shallow trench isolation (STI), which
suitable for embedded applications with low power
consumption. This cell shows good promise for scaling to
next generation.

Reference:
[1]. J. C. Kuo et al., Int'l NVM Tech. Conf. 1998, p.28.
[2] "IEEE Standard Definitions and Characterization of
    Floating Gate Semiconductor Arrays", IEEE Std
    1005-1998
[3] F. Schuler, "Long times Estimation of Failure Rates
    due to Low Temperature Charge Loss", Proc.
    Nonvolatile Semiconductor Memory Workshop,
    (2000)
[4] G. Tempel et al, "Abnormal Charge Loss of Flash
    Cells at Medium Temperatures", Proc Nonvolatile
    Semiconductor Memory Workshop, (2000)
[5] J. Manthey, Ph.D. these (1990)
[6] Cappelletti P., Golla C., Olivo P., Zanoni E. "Flash
    Memories",. , Kluwer Academic 2000

Design Considerations in Scaled SONOS Nonvolatile Memory Devices

Jiankang Bu and Marvin H. White

Sherman Fairchild Laboratory, 16A Memorial Dr. E., Lehigh University, Bethlehem, PA 18015

Phone: (610) 758-4421 Fax: (610) 758-6705 Email: jib2@lehigh.edu and mhw0@lehigh.edu

Abstract — Scaling the programming voltage, while still maintaining 10-year data retention time, has been always a big challenge for Poly-Oxide-Nitride-Oxide-Silicon (SONOS) researchers. We describe our progress in the design and scaling of SONOS nonvolatile memory devices. –9V +10V (1ms) programmable SONOS devices ensuring 10 years retention time after 10^7 Erase/Write cycles at 85°C have been developed successfully. Deuterium anneal, applied in SONOS device fabrication for the first time, improves the endurance characteristics better than traditional hydrogen or forming gas anneal. In this paper, we describe scaling considerations and process optimization along with experiments and characterization results.

TABLE OF CONTENTS

1. INTRODUCTION

Next generation high density electrically erasable programmable read-only memories (EEPROMs) require an endurance in excess of 10^6 erase/write cycles with 10-year data retention at 85°C and low programming voltages 5-10V [1] [2]. Two basic types of EEPROMs exist, namely, the floating gate device and the floating trap device, i.e. SONOS [3-4]. The floating gate device stores charge in the polysilicon as free carriers as a continuous spatial distribution in the conduction band, and the SONOS stores charge in spatially isolated deep level traps (Fig. 1).

The floating-gate memory has been running out of steam with respect to scaling cell-size and program/erase voltages. The relatively thick (7-12 nm) tunnel oxide in the floating-gate type memories provides good 10-year data retention; however, the high voltage requirement [5] has created a reliability issue, as it has exceeded the voltage limits of scaled CMOS devices. Dielectric hot carrier degradation, punch-through avalanche effects and high voltage junction breakdown [6-8] limit the lateral scaling to achieve high density. The concern over the loss of the entire memory charge through a single defect in the tunnel oxide limits vertical scaling and lower programming voltages [2], which increases support circuitry area and reduces the array efficiency.
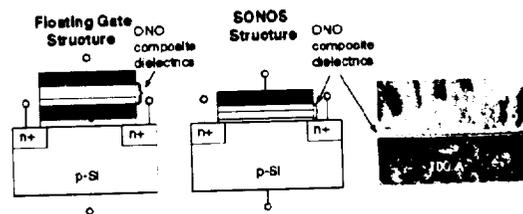
Fig. 1 Floating gate memory and floating trap (SONOS) with device cross-section.

The demand for low-power, low voltage electronics has accelerated the pace for NVSM circuit designers to consider SONOS for low voltage, high density EEPROM's. The motivation for the interest in SONOS lies in low programming voltages, endurance to extended erase/write cycling, resistance to radiation, and compatibility with high density scaled CMOS technology. A 3V, 1Mb, full-featured SONOS EEPROM has been manufactured using 0.8 μm CMOS technology [9]. Fujiwara et al. reported a 0.13 μm SONOS single transistor memory cell [10] with unselected word line bias for program-disturb improvement with subsequent scaling to 0.1 μm and beyond [11].

A radiation-hardened, SONOS based 4K gate field programmable gate array (FPGA) device has been described in 0.8 μm triple-level metal technology [12]. In this array SONOS transistors provide program connectivity and, for the first

9

time, offer the feature of reconfigurability in a FPGA device. SONOS NVSMs are capable of a small cell size ($6F^2$ where F = feature size) [13]. The ultra-thin tunnel oxide can conduct high current via direct tunneling with less charge trapping and a dramatic increase in charge-to-breakdown, Q_{BD}, with tunnel oxide thickness less than 3.2 nm – the mean free path of electrons in the oxide. Thus, we have the possibility of thin tunnel oxide SONOS devices for dynamic/quasi-nonvolatile memory applications, as discussed by Wann et al. [14] and King et al. [15].

A considerable effort has been devoted to scaling the programming voltages of SONOS devices with improved retention characteristics under extended erase/write cycling at elevated temperatures. Minami and Kamigaki reported a SONOS device with 10-year data retention after 10^7 erase/write cycles with programming voltage –11V +13V (1ms) [16]. Reisinger et al. proposed a p+ gate SONOS structure [17] to improve erase speed and data retention time. Libsch et al. [18], French and White [19]. Yang and White [20] have all discussed a 5-8V EEPROM cell for high density NVSM.

Recently, we have explored the scaling of low voltage, long retention SONOS memory devices. SONOS devices fabricated at Lehigh University show a 0.5V detection window at 10-year data retention after 10^7 erase/write cycles at 85°C with programming voltage –9V +10V (1ms). We will describe scaling and process optimization in the Section 2. SONOS device fabrication and characterization results are described in Sections 3 and 4, respectively.

2. SCALING CONSIDERATIONS -- PROCESS OPTIMIZATION--

2.1 ONO Stack Scaling

Three approaches have been described in the literature to obtain good balance between speed, retention and endurance. One approach, taken by Roy and White [21], is to scale the nitride storage layer, but keep the blocking oxide thicker, which increase the memory window (will decrease for scaled nitride layer otherwise) and the amount of charge trapped at nitride/blocking oxide interface is increased too. Another approach, taken by Minami and Kamigaki [15], and Dellin et al. [22], uses a thin blocking oxide just thick enough to block the injection of charge from the gate, and scale the nitride layer at the same time. The other

approach, investigated by Williams et al. [23] and Kapoor et al. [24], employed an oxynitride instead of the nitride as the storage medium, because the oxynitride film has a smaller trap density and hence a smaller Coulombic repulsion between the trapped charges. The tunnel oxide can be further scaled with or without a blocking oxide. In addition, Hu and White [25] have presented a buried channel device instead of a surface channel device to reduce back tunneling.

The work presented in this paper is based on the first approach, namely, incorporating the optimization of the tunnel oxide thickness [26]. For 8-10V program/erase voltages, a 10 nm effective gate dielectric thickness is preferred to guarantee an electric field for modified Fowler-Nordheim tunneling [27]. A 2.0 nm tunnel oxide and thicker blocking oxide (5.5 nm) are used for good retention and reliability considerations. The scaling of the nitride layer is based on the constant tunnel oxide electric field theory, which we will detail in Section 4 with a comparison of different design approaches.

2.2 Process Optimization

In our scaling, a trap-rich, silicon nitride (with a $SiCl_2H_2$: NH_3 ratio of 10:1) is necessary, as silicon nitride films deposited with high $SiCl_2H_2$: NH_3 gas flow ratios show high trap densities, which facilitate fast programming speed [28]. Yang et al. [29] conducted AFM studies on silicon nitride films, which were deposited at different temperatures. Their studies revealed 680°C as the optimum temperature for LPCVD nitride deposition for a minimum surface roughness with improved memory retention. Minami et al. demonstrated LPCVD formed blocking oxides improved data retention dramatically [15]. In addition, tunnel oxides grown at high temperature, exhibit improved performance and reliability [30]. Superior retention and endurance are obtained with the use of a triple-wall oxidation furnace rather than the conventional single-wall furnace [25].

We have performed high temperature (700 °C 4 hours) deuterium anneals instead of hydrogen forming gas anneals after the contact windows are opened. Anneal temperatures comparable to or lower than the nitride deposition temperature provide less migration of the stored charge in the nitride [15]. Also, the interface states generation is reduced under extended program erase cycling and retention reliability is improved as described in Section 4.
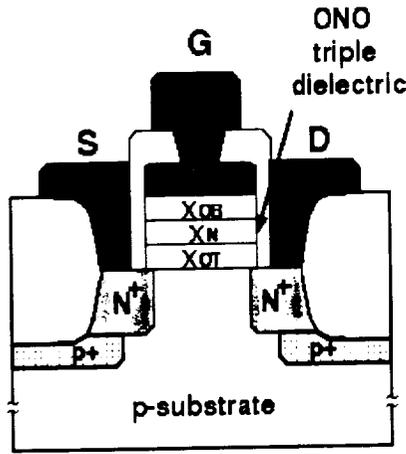
3. SONOS DEVICE FABRICATION

Fig. 2 Device structure of SONOS nonvolatile memory transistor with tunnel oxide 2.0 nm, nitride layer 4.5 nm, and blocking oxide 5.5 nm.

We have fabricated SONOS devices with N-well CMOS technology and LOCOS isolation. The processing sequence is identical to conventional CMOS technology except for the formation of the ONO dielectric stack. The key process steps are as follows: A 2.0 nm thick tunnel oxide is grown at 800°C for 40 min with argon-diluted oxygen (1% O_2 in Ar) in a custom-designed triple wall oxidation furnace followed by 30 min. argon anneal to relieve the stress caused by high temperature oxidation. Next, a 4.5 nm silicon nitride is deposited in a LPCVD reactor for 15 min. at 680°C with gas flow ratio of $SiCl_2H_2$: $NH_3 = 100:10$ (sccm). A 5.5 nm LPCVD blocking oxide is deposited with $SiCl_2H_2$: $N_2O = 10:100$ (sccm) at 725°C followed by a steam densification at 900°C for 30 min. After the ONO triple dielectric film is formed, a layer of polysilicon is deposited and doped in a $POCl_3$ process. Next, the gate is patterned and the contact windows are opened. A 4 hour 10% D_2/N_2 anneal at 700°C is performed to lower the interface state density. This is followed by an aluminum deposition for contact metallization with a post-metal-anneal (PMA) at 400°C for 30 min. in 10% D_2/N_2. For comparison purposes, another group of wafers annealed with 10% H_2/N_2 are fabricated at the same time. Fig. 2 shows the device structure.

4. MEASUREMENT RESULTS AND DISCUSSION

In this section, we present erase/write, retention and endurance electrical characteristics of scaled SONOS devices. The threshold voltage shifts are measured 1 μs after an erase (write) pulse that follows a 10 s low voltage reset pulse of reverse polarity. All measurements are made at 85°C unless specified. We compare the effects of high temperature deuterium and hydrogen anneals.

Fig. 3 shows the erase/write characteristics. The SONOS device can be operated with a 1 ms −9V/+10V pulse. In the +10V write or program operation, the initial electric field across the tunnel oxide is 11.8 MV/cm with 9.8 MV/cm for the −9V erase operation. In previous MONOS/SONOS scaling scenarios the electric field across the tunnel oxide or nitride has been maintained nearly constant while the triple dielectric dimensions are scaled [31-32]. Using the constant tunnel oxide electric field theory as the criteria, we compared devices scaled along different approaches in Fig. 4. For the device with 1.8 nm tunnel oxide reported by Minami et al. [16], the initial electric field at 13V is approximately 10MV/cm, assuming zero stored nitride charge.

If we scale the nitride from 13 nm to 4.5 nm and thicken the tunnel and blocking oxides to maintain the same initial electric field, then we have a 10 nm effective gate dielectric thickness with a programming voltage decreased to 9V. However, a smaller memory window and earlier saturation are observed with our scaled nitride device. A thicker tunnel oxide and blocking oxide compensates for the barrier lowering effect due to Coulombic repulsion in the scaled nitride layer (associated with degraded retention) and offers highly-reliable retention characteristics. French et al. [27] noticed scaling the blocking oxide does not improve the erase/write speed of the device. Fig. 5 illustrates the retention and endurance characteristics of our device. 10-year data retention with 0.5V memory window after 10^7 erase/write cycles at 85°C is ensured.
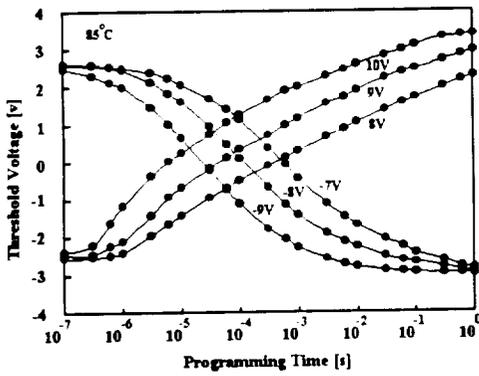
Fig. 3 Erase/write characteristics for scaled SONOS device with tunnel oxide 2.0nm, nitride 4.5nm, blocking oxide 5.5nm.
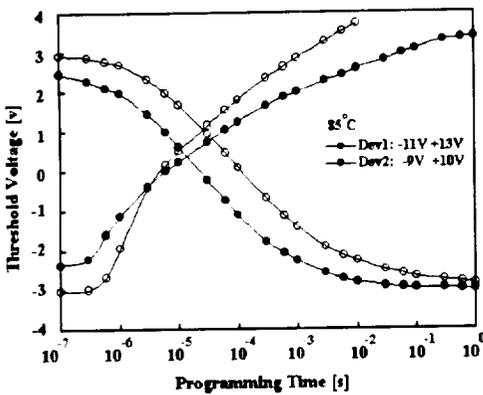

Fig. 4 Comparison of SONOS/MONOS devices with different design approaches. Dev1: tunnel oxide 1.8nm, nitride 13nm, blocking oxide 3nm. Dev2: tunnel oxide 2.0nm, nitride 4.5nm, blocking oxide 5.5nm.
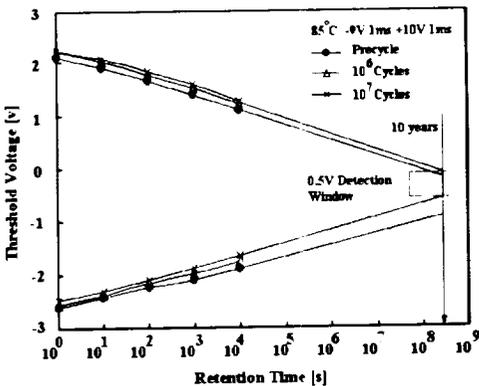

Fig. 5 Retention characteristics after erase/write cycles for SONOS device with tunnel oxide 2.0nm, nitride 4.5nm, blocking oxide 5.5nm.

The deterioration of Si-SiO$_2$ interface is of major concern in NVSM devices because of the high electric fields across the insulators and the continual passage of charge across the tunnel oxide region. This deterioration manifests itself as a buildup of "interface traps", which are defect centers located at the Si-SiO$_2$ interface. Their build-up, along with traps exist between SiO$_2$-Si$_3$N$_4$ layer, is detrimental to both SONOS and floating-gate NVSM operation because they (1) provide an additional shift in the device threshold voltage and (2) degrade long-term retention by increasing the so-called back-tunneling current [33]. Maes et al. have employed high temperature hydrogen anneals to reduce interface trap density and improve data retention time [34].

The channel hot carrier lifetime of MOSFETs, annealed in a deuterium ambient instead of the traditional hydrogen or forming gas ambient, have increased by an order of magnitude [35]. In an extension of these studies, we have examined high temperature deuterium anneals in the fabrication of SONOS devices. Fig. 6 compares the interface trap densities (D$_{it}$) of SONOS devices annealed in deuterium and hydrogen environments. The initial D$_{it}$ is nearly the same for both devices. However, under extensive erase/write cycling, more interface traps are created for hydrogen annealed devices than with deuterium annealed devices.

The retention characteristics of hydrogen annealed SONOS devices are shown in Fig. 7. In contrast with Fig. 5, deuterium annealed SONOS devices have nearly an order of magnitude longer retention time after 10^7 erase/write cycles at 85°C than hydrogen annealed devices for the same detection window. Studies have been conducted to investigate this isotopic interfacial hardening effect [36]. These studies suggest the improved robustness of interface states to dissociation is associated with the difference in vibration mode frequencies of the Si-H and Si-D configurations. The vibration frequency of the bending mode for Si-D bonds is around 460 cm^{-1}, which is very close to the frequency of one of the bulk silicon phonon modes (463 cm^{-1}). Coupling of these modes can provide an energy relaxation channel and make the dissociation of Si-D bonds more difficult than the dissociation of Si-H bonds.
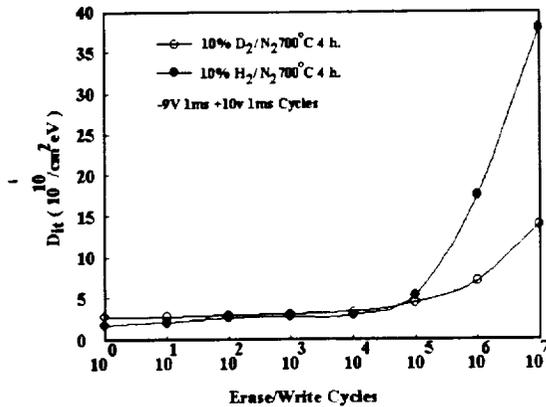
12

Fig. 6 Dit variation of high temperature deuterium annealed and hydrogen annealed devices.
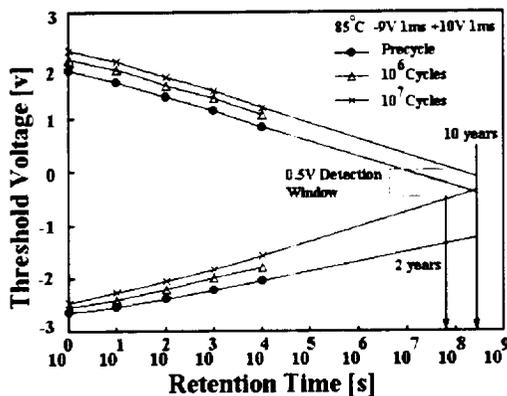

Fig. 7 Retention characteristics of a hydrogen-annealed SONOS device. Compared with deuterium-annealed SONOS in Fig. 5, retention time is nearly one order of magnitude shorter for the same detection window.

5. CONCLUSIONS

SONOS nonvolatile memory devices exhibit a 0.5V detection window with 10-year data retention after 10^7 erase/write cycles at 85°C, with 1 ms +10V/-9V program/erase voltages. Deuterium annealing offers improved endurance characteristics over traditional hydrogen forming gas anneals. These SONOS devices are promising candidates for low voltage, radiation-hardened, high-density EEPROM's applications.

6. ACKNOWLEDGMENTS

7. REFERENCES

1. M. H. White and D. Adams, "Low-Voltage SONOS Nonvolatile Semiconductor Memories (NVSMs)," GOMAC 2000.

2. Y. Kamigaki, S. Minami, T. Hagiwara, K. Furusawa, T. Furuno, K. Uchida, M. Terasawa, and K. Yamazaki, "Yield and reliability of MNOS EEPROM products," IEEE J. Solid-State Circuits, vol. 24, p. 1714-1722, Dec. 1989.

3. J. Chang, "Nonvolatile Semiconductor Memory Devices," Proc. IEEE, 64(7), p. 1039-1059, July 1976.

4. F. R. Libsch and M. H. White, SONOS NVSM, Chpt. 5, Nonvolatile Semiconductor Memory Technology, ed. by W. D. Brown and J.E. Brewer, IEEE Press Series on Microelectronic Systems, p. 309-357 (1997).

5. N. Ajika, M. Ohi, H. Arima, T. Matsukawa, N. Tsubouchi, "A 5 volt only 16M bit FLASH EEPROM cell with a simple stacked gate structure," in IEDM Tech. Dig., pp. 115-118, 1990.

6. S. Asia, Proc. IEEE 74, pp. 1623, 1986.

7. P.C.Y. Chen, IEEE Trans. Elect. Dev. ED-24, 1977.

8. E. Takeda, Symp. VLSI Tech., p. 2, Dig. Tech. Papers, Kobe, Japan, 1985.

9. S. Minami, K. Ujiie, M. Terasawa, K. Komori, K. Furusawa, Y. Kamigai, "A 3 volt 1 Mbit full-featured EEPROM using a highly-reliable MONOS device technology," IEICE Trans. Elect. Vol. E77-C pp. 1260-1269, 1994.

10. I. Fujiwara, "A 0.13 μm MONOS single transistor memory cell with separated source lines," Proc. IEDM, San Francisco, CA 1998

11. I. Fujiwara H. Aozasa, A. Nakamura, Y. Hayashi and T. Kobayashi, "MONOS memory cell scalable to 0.1 μm and beyond," Proc. 2000 NVSM Workshop, Monterey, CA p. 117.

12. D. Mavis et al., "A reconfigurable nonvolatile radiation hardened FPGA for

space applications," *Proc. 1998 NASA MAPLD Conf.*

13. M. H. White and Y. Yang, "A low voltage SONOS nonvolatile semiconductor memory technology," *IEEE Trans. Comp., Pkg and Mfg. Tech.,* vol. 20, p.190 1997.

14. H. C. Wann and C. Hu, "High-endurance ultra-thin tunnel oxide in MONOS device structure for dynamic memory application," *IEEE Elect. Dev. L.,* vol. 16, pp. 491, 1995.

15. Y. King, T. King, C. Hu, "A long-refresh dynamic/quasi-nonvolatile memory device with 2 nm tunneling oxide," *IEEE Elect. Dev. L.,* vol. 20, no. 8, pp. 409, 1999.

16. S. Minami, Y. Kamigaki, "A novel MONOS nonvolatile memory device ensuring 10-year data retention after 10^7 erase/write cycles," *IEEE Trans. Elect. Dev.,* vol. 40, 1993.

17. H. Reisinger, M. Franosch, B. Hasler and T. Bohm, "A novel SONOS structure for nonvolatile memories with improved data retention," *Symp. VLSI Tech.* p. 113, 1997

18. F. R. Libsch, A. Roy, and M. H. White, "A true 5V EEPROM cell for high density NVSM," *IEEE Trans. Elect. Dev.,* vol 34, pp 2372, 1987.

19. M. L. French and M. H. White, "Scaling of multidielectric nonvolatile SONOS Memory Structures", *Solid-State Elect.,* vol. 37, p. 1913, 1995.

20. Y. Yang and M. H. White, "Charge Retention of Scaled SONOS Nonvolatile Memory Devices at Elevated Temperatures", *Solid-State Elect.,* August 2000.

21. A. Roy, F. R. French, and M. H. White, "Investigations on ultrathin silicon nitride and silicon dioxide films in nonvolatile semiconductor memory transistors," *Proc. Symp. Silicon Nitride and Silicon Dioxide Thin Insulating Films,* San Diego, CA, 1986.

22. T. A. Dellin and J. P. McWhorter, "Scaling MONOS nonvolatile memory transistors," *Proc. Symp. Silicon Nitride and Silicon Dioxide Thin Insulating Films,* San Diego, CA, 1986.

23. D. Williams, D. Adams, R. Bishop, "Radiation hardened 64K/256K EEPROM technology," *Proc. Int'l Nonvolatile Memory Tech. Conf.* pp. 67, 1996.

24. V. J. Kapoor, R. S. Bailey, and R. A. Turi, "Chemical composition, charge trapping, and memory properties of oxynitride films for MNOS devices," *J. Electrochem. Soc.,* vol. 137, 1990.

25. Y. Hu and M. H. White, "Charge retention in scaled SONOS nonvolatile semiconductor memory devices – modeling and characterization," *Solid State Elect.,* vol. 36, p. 1401, 1993.

26. S. Minami and Y. Kamigaki, "Tunnel oxide thickness optimization for high-performance MNOS nonvolatile memory devices," *IEICE Trans. Electron.,* vol. E74, pp.875, 1991.

27. M. L. French, M. Chen, H. Sathianathan, M. H. White, " Design and scaling of a SONOS multidielectric device," *IEEE Trans. Comp. Pkg. and Mfg. Tech., part A,* vol. 17, 1994.

28. S. Fujita, H. Toyoshima, M. Nishihara and A. Sasaki, "Variation of trap states and dangling bonds in CVD Si_3N_4 layer on Si substrate by NH_3/SiH_4 ratio," *J. Elec. Mat.,* vol. (11), no.4, pp.795, 1982.

29. Y. Yang, A. Purwar, M. H. White, "Reliability considerations in scaled SONOS nonvolatile memory devices," *Solid State Elect.,* vol. 43, pp. 2025-2032, 1999.

30. A. B. Joshi, G. Q. Lo, D. K. Kwong, "Improved performance and reliability of MOSFETs with thin gate oxides grown at high temperature," 29th *Proc. IRPS,* pp. 316, 1991.

31. F. R. Libsch and M. H. White, "Charge transport and storage of low programming voltage SONOS/MONOS memory devices," *Solid-State Elect.,* vol. 33, pg. 105, 1990.

32. S. Minami, Y. Kamigaki, "New scaling guidelines for MNOS nonvolatile memory devices," *IEEE Trans. Elect. Dev.,* vol. 38, pp. 2519, 1991.

33. M. H. White, D. Adams and J. Bu, "On the go with SONOS," *IEEE Circuits and Dev.,* vol. 16, no. 4, 2000.

34. H. E. Maes, S. H. Usmani, and G. L. Heyns, "Effects of a high temperature hydrogen anneal on the memory retention of metal-nitride-oxide-silicon transistors at elevated temperatures," *J. Appl. Phys.,* vol. 52, pp. 4348, 1981

35. J. W. Lyding and K. Hess, "Reduction of hot electron degradation in metal oxide semiconductor transistors by deuterium processing," *Appl. Phys. L.,* vol. 68, p. 2526, 1996.

36. H. C. Mogul, L. Cong, R. M. Wallace, P. J. Chen, T. A. Rost and K. Harvey, "Electrical and physical characterization of deuterium sinter on submicron devices," *Appl. Phys. L.,* vol. 72, pp. 1721, 1998.

14

8. BIOGRAPHIES

Jiankang Bu was born in Hebei Province, China on February 4, 1972. He received his B.S. degree (1994) and M.S. degree (1997) in Electrical Engineering from Nankai University, China. He joined the Electrical Engineering Dept. at Lehigh University in the Fall 1997 as a research assistant. He is currently working on his Ph.D degree in Electrical Engineering on measurement circuit design, device characterization and fabrication of scaled SONOS nonvolatile memories. He is a member of the IEEE Electron Devices Society (EDS) and the Sigma Xi research honorary.

Marvin H. White was born in the Bronx, New York on September 6, 1937. He received an A.S. degree in Engineering from the Henry Ford Community College (1957) a B.S.E. degree in Physics and Math (1960), M.S. degree in Physics (1961) from the University of Michigan and a Ph.D. degree in Electrical Engineering (1969) from the Ohio State University. In 1961 he joined the Westinghouse Solid-State Laboratory in Baltimore, MD. where he worked on advanced military and NASA imaging systems. From 1961 - 1981 he worked at Westinghouse as an Advisory Engineer in the design of low-power, custom integrated circuits with technologies of CMOS, Bipolar, MNOS and CCDs. During this period he was an adjunct Professor at the Electrical Engineering Depart-ment of the University of Maryland and a visiting Fulbright Professor at the Catholique Universite' de Louvain in Louvain-la-Neuve, Belgium.

In 1981, he became the Sherman Fairchild Professor in Solid-State Studies and Electrical Engineering at Lehigh University. At Lehigh he has developed a graduate program in microelectronics with research on SONOS nonvolatile memory devices, CMOS device modeling, studies of the Si-SiO$_2$ interface, SiC devices, and custom integrated circuits and sensors. He has graduated 23 Ph.D. students in microelectronics. He has served as a Visiting Researcher at the Naval Research Laboratories (1987) and a Program Director in Solid- State and Microstructures at the National Science Foundation (1995-96). In 1997 he received the Eleanor and Joseph Libsch Research Award at Lehigh University. He is currently the Director of the Sherman Fairchild Center for Solid-State Studies.

Prof. White is an IEEE Fellow (1974) and the recipient of the J. J. Ebers Award (1997) and the Masaru Ibuka IEEE Consumer Electronics Award (2000). In 1982 he was the IEEE Electron Devices Society (EDS) National Lecturer and is presently a Distinguished EDS Lecturer. He has served on IEEE/EDS committees, in particular, membership, and education. He is a member of Eta Kappa Nu and Sigma Xi.

3-D Computer Simulation of Nano-crystal Floating –Gate Flash Memory Devices

Aaron Thean and Jean-Pierre Leburton
Beckman Institute for Advanced Science and Technology.
University of Illinois, Urbana, Illinois 61801.
vthean@uiuc.edu

Michael Sadd
Motorola Inc, Semiconductor Products Sector
3501 Ed Bluestein Blvd., MD-K20
Austin, Texas 78721.

Abstract—We present a detailed investigation of single-electron charging effects in silicon nano-crystal quantum- dots, taking into account the quantum-mechanical properties of the silicon bandstructure. We show that the retention and erase times resulting from differences in the quantum coupling between the channel states and the nano-crystal states of different geometries can differ by orders of magnitude.

1. INTRODUCTION

Metal-oxide-semiconductor(MOS) field-effect-transistor devices with an embedded granular layer of nano-crystals in the dielectric as a floating gate have generated much interest as a promising ultra-low-power, high-endurance and ultra-fast non-volatile memory option. These devices have so far demonstrated superior data retention, write/erase endurance characteristics and short write times, compared to conventional floating-gate flash memories[1]. The improved data retention property results from the strong confinement of charge stored in the nano-crystals which limits the draining of the floating-gate charge due to oxide traps and lateral leakages to the source/drain contact regions. With the charge leakage minimized, a thinner dielectric can be used as a tunnel barrier to separate the nano-crystal from the channel. Consequently, the direct quantum-mechanical tunneling of electrons or holes through the thin tunnel oxide results in dramatically lowered write and erase voltages, thereby giving access to the low-power and high endurance capability of this technology. Moreover, the possibility of producing nano-crystals of different sizes and shapes, ranging from hemisphere to sphere, has already been demonstrated [2],[3]. This gives rise to the prospect of engineering the floating-gate device performance through the control of nano-crystal size, shape and density. In this paper we present 3-D computer simulations based on the solution of the Schrödinger and Poisson equations to investigate the physics of charging and tunneling process in Si nano-crystal floating-gate flash memory devices.

2. MODEL

Fig. 1 shows the MOS device structure with nano-crystals that are shaped like a truncated-spheres. For the sake of simplicity, we focus on one Si quantum-dot(QD) embedded in the SiO_2 dielectric between a control gate and a p-doped silicon substrate.
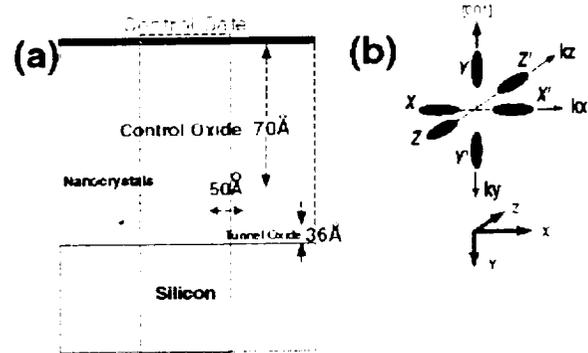
Fig.1. (a) Schematic of the simulated unit cell of the Metal-Oxide-Semiconductor(MOS) structure with a silicon nano-crystal quantum-dot(QD) embedded in the SiO_2. (b) Illustration showing the orientation of the nanocrystal QD and the the six doubly-degenerate conduction band valleys considered in our model. The valleys are labeled with respect to the orientation of their principle axis in k-space.

We assume that the nano-crystals are orientated along the Si[001] direction(Fig. 1(b)), normal to the Si/SiO_2 tunnel oxide interface and the electrons occupy only the lowest sixfold-degenerate X-valleys. We have labeled the six conduction band valleys in k-space according to the orientation of the principal axes, i.e. the two equivalent valleys along the k_x-direction are labeled as X-X', and similarly, Y-Y', Z-Z', for the valleys along k_y and k_z, respectively.

16

The 3-D potential, $\phi(r)$ within the device is obtained from the solution of the nonlinear Poisson equation,

$$\nabla\varepsilon(r)\nabla\phi(r) = -q \ [\ p(r)-n(r)-N_A^-(r) \ ], \quad (1)$$

where $N_A^-(r)$ is the ionized acceptor density, which is chosen to be 3×10^{17} cm^{-3}; the other parts of the device structure remain undoped. Variation of the dielectric permitivity, $\varepsilon(r)$, across the Si/SiO$_2$ interface is also taken into account. We model the control gate of the structure as a metallic contact, which imposes a Dirichlet boundary condition on the electrostatic potential at the top of the structure, i.e. $\phi = \phi_m + V_G$, where ϕ_m is the Femi-level pinning due to the flat-band voltage and V_G is the applied bias. The substrate is chosen to be thick enough($>2\mu$m) so that a zero-electric-field Von-Neumann boundary condition is imposed at the bottom of the structure. Since the QDs are well-isolated from one another, we assumed that the randomness of the QD position in the dielectric layer have little or no influence on the electrostatic interactions between the QDs. Therefore, we impose a periodic boundary condition on the potential, on the sides of the unit-cell. This periodic boundary condition should be valid for even relatively small transistors, as long as they contain a high-density of QDs. This is due to the fact that the transistor behavior should be mainly influenced by the QDs over the central active region of the channel. Moreover, fringing fields from the device edges only affect a small number of QDs in these regions. Since the QDs should be well-separated, the rest of the QDs should remain largely unaffected.

The electronic energy levels within the QD are obtained by solving three 3-D eigen-value Kohn-Sham equations, corresponding to the three two-fold-degenerate valley bandstructure(ν=X,Y,Z). The possibility of intervalley splitting is not considered here.

$$\nabla M_\nu^{-1}\nabla \psi_\nu(r) + \{ -q\phi(r) + \Delta E_C(r) + \varepsilon_{xc}(n(r)) \} \psi_\nu(r)$$
$$= E_\nu \psi_\nu(r) \quad (2)$$

In Eq. 2, $\Delta E_c(r)$ is the material band offsets between the silicon nanocrystals and the silicon dioxide insulator which provides the hard-wall confinement in the QD. M_ν is the electron effective mass tensor that describes different effective masses arising from the anisotropy of the Si bandstructure. The boundary conditions for the Schrödinger equations are imposed such that the bound - state wavefunctions decay to zero before they reach the device boundaries. The electron-electron interaction in nanocrystal QDs is treated within the density functional theory, with the exchange-correlation potential, $\varepsilon_{xc}(n)$, computed from the local density approximation(LDA)[4].

In the nano-crystal, the hole density, $p(r)$ is set to zero and the electron density is given by $n_\nu(r) = \Sigma_\nu \ g_\nu \ |\psi_\nu|^2$,

where g_ν is the electron state occupation. Otherwise, in bulk silicon, $n(r)$ and $p(r)$ are computed within the Thomas-Fermi Approximation. In this analysis, we do not resolve individual electron spins.

3. DISCUSSION

3.1 Stark Effect

Fig. 2 shows the variation of the single-particle energy spectrum of an empty semi-spherical QD(Fig. 1), with gate bias, for the first six levels in a large QD(d=125Å)(Fig.2(a)) and a small QD(d=70 Å)(Fig. 2(b)). The energy levels that arises from the six conduction band valleys form three spin-degenerate sets. The levels are labeled in the form $(n_x,n_y,n_z)_i$, where the i index corresponds to the label for the pair of degenerate valleys, described in Fig. 1(b). The (n_x,n_y,n_z) indices describes the number of nodes of the wavefunction, corresponding to a energy level, in the x,y and z directions, respectively. The lowest energy levels correspond to that of the $(0,0,0)_{YY'}$ states, which are derived from valleys Y and Y', where the principal symmetry axis is along the y-axis. These states are s-like with no nodes. The y-oriented semi-spherical hard wall structure of the QD and the y-directed electric field from the control gate creates the strongest confinement of electrons in the y-direction. Since, electrons in valleys Y and Y' have the heaviest effective mass oriented along the y-axis, the states corresponding to these valleys are the lowest in the energy spectrum.

Close examination of the spectrum in Fig. 2(a) shows a peculiar effect where the order of the energy levels depends strongly on the gate bias. As the device is in the strong accumulation mode($V_G < -3V$), we see that the next upper states, also arise from valleys Y and Y'$((1,0,0)_{YY'}$ and $(0,0,1)_{YY'})$, and are four-fold-degenerate single-node p-like states. However, as the bias becomes more positive, the valley Y and Y' states are gradually displaced by the four-fold--degenerate s-like states from the four different valleys(valleys X,X',Z,Z'). At some point of the gate bias, the s-like states from the X and Z-valleys actually becomes degenerate with the p-like states from the Y-valleys. Comparing the energy spectrum of the larger QD (Fig. 2(a)) with that of the smaller QD(Fig. 2(b)), we see that energy-crossings occur for different states and at different bias points. Moreover, the order of the states in the 70 Å-QD are different from the 125 Å-QD as well. This effect results from the subtle interplay between the effects of anisotropic electronic effective mass, the confining potential of QD and the applied electric field. Moreover, these effects are a strong function of the nano-crystal size.
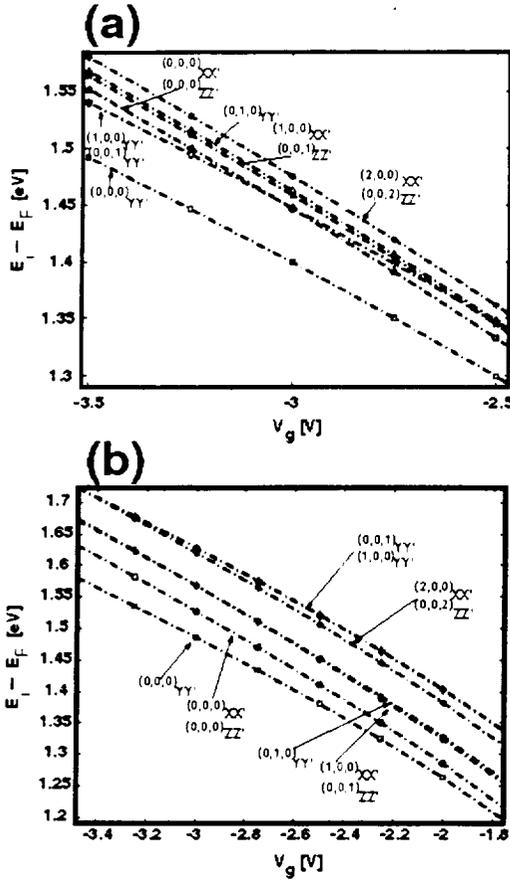
17

(a)

(b)

Fig. 2 (a) Variation of the first six electron energy levels in a empty truncated-sphere nano-crystal of diameter 125Å as a function of the gate bias during strong accumulation in the channel of the MOS structure. (b) Energy levels in a 70 Å-diameter truncated-sphere nano-crystal.

The applied gate electric field produces additional electrostatic confinement within the nano-crystal, in the direction normal to the tunnel oxide(Stark effect). Fig. 3 shows a 2-D slice of the 3-D ground-state wavefunction$((0,0,0)_{YY'})$ through the QD center. Superimposed on the same figures are the contour plots of the potential and the electron density.
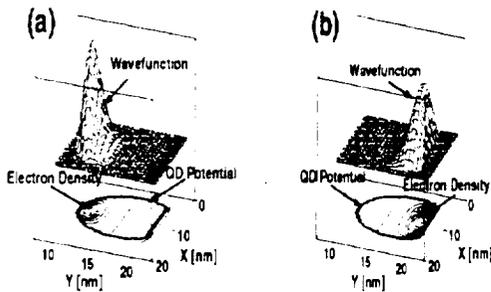

Fig. 3 (a) 2-D slice of the groundstate wavefunction in the nanocrystal for V_G = 1.0V. Superimposed on the same figure are the contour plots of the electron density corresponding to one electron and the confining potential of the nanocrystal. (b) The wavefunction and contour plots of the electron density and potential for V_G = -3.0V.

The Stark effect causes the low-energy electron wavefunctions to be localized near the top of the QD structure, away from the tunnel oxide(Fig. 3 (a)), when a positive gate bias is applied(inversion).

This suggests that charge leakage through these states would be curtailed when strong positive gate bias is maintained. On the other hand, the electron wavefunction moves to the bottom of the nanocrystal(Fig. 3 (b)) when a negative gate bias is applied to discharge the QD floating gate(accumulation). The plot of he wavefunction for channel inversion(Fig. 3(a)) shows that the electrons are confined to a tighter region at the QD top where, the electronic states are more sensitive to the influence of the hard-wall potential than to the gate electric field.

3.2 Single-Electron Charging

Single-electron charging is simulated only in the low temperature limit(T=25mK). This enables us to determine the integer number of electrons, N, in the QD, with the use of the Slater formula[5][6] which gives directly the difference between the total free energy, E_T, for the two occupations of N and N+1 electrons at each gate bias V_G.

$$E_T(N+1) - E_T(N) = \int_0^1 \varepsilon_{LAO}(n)dn \; \square \; \varepsilon_{LAO}(\frac{1}{2}) - E_F \quad (3)$$

In Eq. 3, ε_{LAO} corresponds to the eigenvalue of the lowest available orbital computed from the Kohn-Sham equation occupied by half an electron. E_F is the Fermi energy. Hence, if $\varepsilon_{LAO}(1/2)$ - E_F is positive then $E_T(N+1)$- $E_T(N)>0$ and the QD contains N electrons, otherwise it contains N+1 electrons.

Fig. 4 shows the coulomb staircase that describes the number of electrons in the truncated-sphere QD as a function of the control gate bias. The charging of the QD with the first electron occurs when the empty ground state$((0,0,0)_{YY'})$ crosses the Fermi level and this crossing point depends on the flat-band voltage of the device. As such the step corresponding to the charging from N=0 to N=1 is large. The somewhat uniform steps from N=1 to N=4 results from similar coulombic interaction between the four-fold degenerate states from valleys Y and Y'. The addition of the fourth electron fills the first s-shell. Therefore, the fifth electron goes into states belonging to valley X, $(0,0,0)_X$, and gives rise to an increase in step size. This is represented by a peak in the addition energy at N=4 as shown in Fig. 4(inset). We note that the average addition energy for the nano-crystal QD is of the order of a tenth of an electron volt.
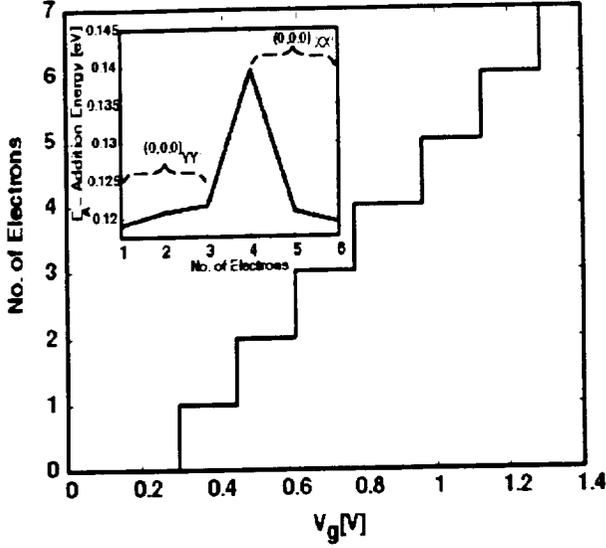
18

Fig. 4 Coulomb staircase of the 125Å nanocrystal at T=25mK showing the variation of the electron number in the nanocrystal QD as a function of the control gate voltage. (Inset): Addition energies of the nanocrystal as a function of the number of electrons in the QD.

3.3 Retention Time

To estimate the retention time of the electron in the nano-crystal, we solve for the quasi-bound states of the QD by imposing a quantum open boundary between the QD and the channel for Eq.(2)(Fig. 5).
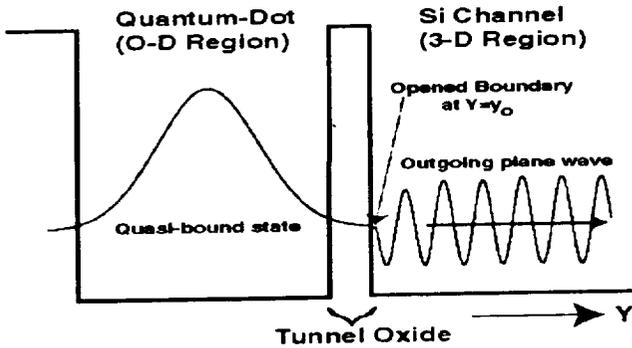

Fig. 5 A schematic illustrating how a quasi-bound state in the quantum-dot(0-D region) is coupled to an outgoing plane wave in the silicon channel(3-D region) through an opened boundary just beyond the tunnel oxide.

We assume that the diverging spherical components of the outgoing wave in the 3-D region are weak for a thick tunnel oxide. Hence, the wavefunction in the channel region is approximated by a plane wave propagating in the positive y-direction. Numerically, the opened boundary is implemented by coupling the discretized 0D wavefunction to an outgoing plane wave, point-by-point, across the tunnel oxide-silicon channel interface at $Y=y_0$

(4) and ensuring the conservation flux across the same interface(6).

$$\psi(x, z; y - y) = \psi(x, z; y) e^{ik \cdots}$$ (4)

$$k = \frac{\sqrt{2m_{Si}(E_r - E_r)}}{\hbar}$$ (5)

$$\frac{1}{m_{SiO_2}} \frac{\partial \psi(x, y, z)}{\partial y}\bigg|_{y_0-} = \frac{1}{m_{Si}} \frac{\partial \psi(x, y, z)}{\partial y}\bigg|_{y_0+}$$ (6)

In Eqs. 5 and 6, m_{Si} and m_{SiO_2} are the electron effective masses in the silicon and silicon dioxide regions, respectively. The wavevector k, in (4) is a function of the propagating wave energy in the channel region. This energy is given by the difference between the eigen-energy of the quasi-bound state(E_r) and the conduction band-edge of in channel region(E_c)(5). This means that the solution of (2) becomes a nonlinear iterative problem where the opened boundary condition depends self-consistently on the eigen-energy solution.

By coupling the QD to the opened channel region, the Hamiltonian described by (2) is no longer a Hermitian operator and the eigen-energies of the quasi-bound states are now complex. The imaginary component of the eigen-energy is related to the decay lifetime of the eigen-state(7). Therefore, we choose to define the retention time of the QD by this decay time-constant, τ.

$$\tau = \frac{\hbar}{2 \, \text{Img}(E_r)}$$ (7)

In this paper we choose to compare the influence of the QD geometry on the retention time. We consider specifically three different shapes – hemisphere, truncated sphere and sphere(Fig. 6)
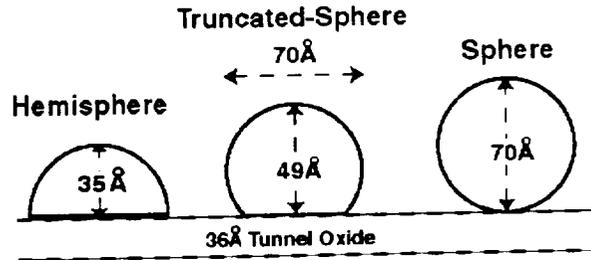
Fig. 6 Illustration showing the three differently-shaped nano-crystals on the tunnel oxide.

We assume that under the flat-band condition of the MOS structure the electron resides mostly in the highly-degenerate ground-state. Hence, retention time for the zero-electric-field condition of the nano-crystal is dominated by the ground-state lifetime. Table I shows the retention time comparison between the hemispherical, the truncated-sphere and the spherical QDs. We compare only the lifetimes of the ground-state for the different nano-

19

crystal geometries. We found that the retention time can vary dramatically from days to years depending on the geometry of the QD. The electron tunneling process is highly sensitive to both the confining volume of the nanocrystal and the tunneling cross-section. The hemispherical geometry, with the smallest volume, gives rise to a ground-state which is at a higher energy compared to both the truncated-sphere and spherical cases. Moreover , the effective tunneling cross-section is also the greatest for the hemisphere. These two factors contribute to its much reduced retention time. Since, the lifetime of the quantum state increases with decreasing energy and decreasing tunnel cross-sectional area , the retention time increases for the truncated-sphere and the sphere.

Shapes	Retention Time
Hemisphere	11 Days
Truncated Sphere	3 Months
Sphere	10 Years

Table I. Retention times for a 7nm-diameter silicon nano-crystals of on a 3.5nm thick tunnel oxide. under flat-band condition of the device.
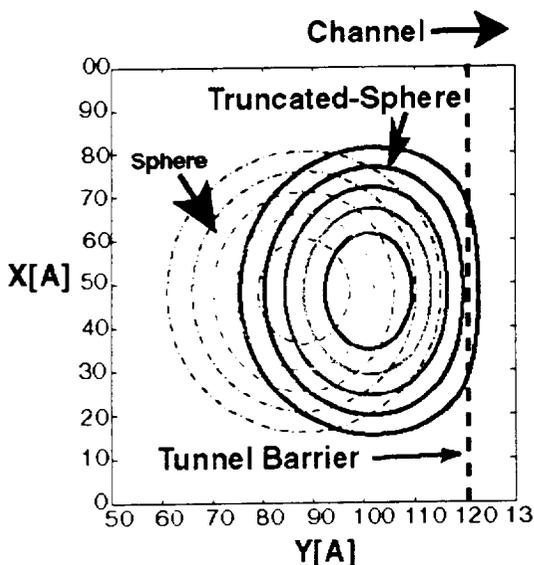

Fig. 7. Contour plots showing the real components of the ground-state wavefunctions for the truncated-sphere and the spherical nanocrystals

In Fig. 7, we show the geometrical effect on the different degrees of quantum-mechanical coupling between the nanocrystal states and unbound channel. The wavefunction of the truncated-sphere(solid line) has a greater extension into the tunnel barrier, compared to the wavefunction of the sphere(dashed lines). This means that there is a stronger coupling between the QD wavefunction and the channel states for the hemisphere, resulting a

larger effective tunneling cross-section for the truncated-sphere.

4. SUMMARY

We have shown that depending on the nano-crystal size, the effects of electrostatic confinement due to the control-gate and the hard-wall potential in the nano-crystals can strongly influence the ordering of the electronic states. In addition, the retention and erase times vary dramatically depending on the shape of the nano-crystals.

5. ACKNOWLEDGMENTS

REFERENCES

[1] S. Tiwari et. al., Tech. Dig. Of IEDM, p.512 (1995)
[2] S. Tiwari et. al., Appl. Phys. Lett. 68, p.1377 (1996)
[3] A. Nakajima et. al. J. Appl. Phys. 84 , p 1316 (1998)
[4] J. Perdew and A. Zunger, Phys. Rev. B, 13, p. 556 (1980)
[5] J.C. Slater, Adv. Quantum Chem. 6, 1 (1972).
[6] J.F. Janak, Phys. Rev. B 18, 7165 (1978).

BIOGRAPHIES

Aaron Thean received his B.S.E.E and M.S. degrees from the University of Illinois at Champaign-Urbana in 1996 and 1997, respectively. He is currently with the Computational Electronics Group at the University of Illinois pursuing his Ph.D. in Electrical Engineering.

Michael Sadd received the B.A. degree in physics in 1991 from Grinnell College, Grinnell, IA. He earned his M.S. and Ph.D. degrees in physics in 1996 and 1998 from Cornell University, Ithaca, N.Y. Since 1998, he has been with the Motorola's Digital DNA Laboratories in Austin, Texas.

Jean-Pierre Leburton received his Ph.D. in Physics from the University of Liege (Belgium) in 1978. In 1979, he was a research scientist with Siemens A.G. in Munich, Germany, and joined the University of Illinois in 1981 where he is currently a Professor in the Department of Electrical and Computer Engineering, and a full-time member in the Beckman Institute. In 1992 and 2000, he was a Visiting Professor respectively in the Research Center for Advanced Science and Technology(RCAST) at the University of Tokyo, and in the Federal Polytechnic Institute (EPFL) in Lausanne, Switzerland. Professor Leburton' s fields of professional interest are nanoscale semiconductor devices, hot electrons and quantum transport in nanostructures, electronic and optical properties of quantum well, superlattices, and low-dimensional structures. His approaches involve use of advanced numerical techniques such as Monte Carlo simulation and 3D self-consistent Schroedinger-Poisson models. He has been chairman of several

international conferences and symposia, and published more than 200 papers in technical journals. Professor Leburton is "Chevalier dans L'Ordre des Palmes Academiques" (French Government), Fellow of the IEEE and Fellow of the American Physical Society, and Associate to the Center for Advanced Study of the University of Illinois.

SESSION #2: Challenges and Applications

Wednesday, November 15, 2000
11:00 AM–3:40 PM
Chair: *Patricia MacFarlane*, Naval Research Laboratory,
Washington D.C.

11:00 AM	Challenges in Implementing Commercial Non-Volatile Memory in Spacecraft Solid State Recorders; *L. Adams, K. Chao, M. Fehringer, C. Miller and P. Murray*, SEAKR Engineering
11:20 AM	Expectations of MRAM in comparison with other non-volatile memory technologies; *K.M.H. Lenssen, G.J.M. Dormans and R. Cuppens*, Philips Research Laboratories and Philips Semiconductors
11:40 AM	Opportunities and Challenges for Embedded Flash Memory; *S. Kianian and D. Sweetman*, Silicon Storage Technology
12:00 PM	LUNCH
1:00 PM	Invited Talk
1:40 PM	Radiation Issues and Applications of Floating Gate Memories; *L. Z. Scheick and D. N. Nguyen*, Jet Propulsion Laboratory
2:00 PM	Total Dose Radiation Response and High Temperature Imprint Characteristics of Chalcogenide Based RAM Resistor Elements; *S. Bernacki*, Raytheon Systems, *K. Hunt and S. Tyson*, AFRL, *S. Hudgens, B. Pashmakov and W. Czubatyj*, Ovonyx
2:20 PM	SONOS Non-Volatile Semiconductor Memories for Space and Military Applications; *D. Adams*, Northrop Grumman, *D. Mavis*, Mission Research Corporation, *J. Murray*, Sandia National Laboratories, *M. White*, Lehigh University
2:40 PM	SONOS Non-Volatile Shadow RAMs for Space Applications; *G. Derbenwick, D. Kamp and A. Isaacson*, Celis Semiconductor, *J. Gill and S. Linn*, Simtek
3:00 PM	Implementation of Ferroelectric Memories for Space Applications; *S. Philpy, G. Derbenwick, D. Kamp and A. Isaacson*, Celis Semiconductor
3:20 PM	(Break)
Late Paper	Ferroelectric Non-Volatile Memories (FRAM); *G. Fox and T. Davenport*, Ramtron International

Challenges in Implementing Commercial Non-Volatile Memory in Spacecraft Solid State Recorders

Lynn Adams (ladams@seakr.com), Kedong Chao (kedong@seakr.com), Matt Fehringer (matt@seakr.com), Chris Miller (chrism@seakr.com), Paul Murray (paul@seakr.com)

SEAKR Engineering, Inc.
12847 East Peakview Ave.
Englewood, CO
(303) 790-8499

Abstract— SEAKR Engineering produces a variety of Solid State Data Storage and Processing Systems for use in applications from avionics to deep space. Power requirements for deep space missions require the use of non-volatile storage system. For the JPL X2000 program, SEAKR selected FLASH memory for use in the PCI Non-Volatile Memory Slice (NVMS). While FLASH Memory offers a significant power advantage for interplanetary missions, there are several constraints intrinsic to FLASH memory devices that must be resolved by the system design. These features include; radiation sensitivity, slow read/write speed, bad memory blocks, limited write cycles, and electro-magnetic noise. This paper describes how these features were resolved and the tradeoffs that were made to accommodate the use of FLASH memory in an interplanetary environment.

TABLE OF CONTENTS

1. INTRODUCTION

SEAKR Engineering provides solid state memory systems, digital signal processors, and memory boards to the international space and avionics communities. SEAKR Engineering now has over 21 successful on-orbit memory systems supporting missions ranging from low Earth orbit to interplanetary spacecraft.

The interplanetary environment presents special challenges to the use of solid state memory. Of primary concern is limited power availability. Limited power drives the need for the use of a non-volatile system as a non-volatile system uses no power unless it is undergoing a read/erase/write cycle. Environmental radiation is the second major concern. Since interplanetary spacecraft are not shielded by the earth's magnetic field, the memory system must survive a high level of natural radiation. The capabilities of memory devices must be evaluated under these conditions to ensure a successful mission.

SEAKR Engineering is working with JPL's X2000 program to developing a non-volatile PCI memory card, the Non-Volatile Memory Slice (NVMS), for use on multiple interplanetary missions.

The SEAKR Engineering NVMS FLASH Memory Card is a multi-capacity memory storage module conforming to the Compact PCI architecture. The NVMS card consists of up to eight memory banks, with each containing up to 256 MB of memory.

2. NON-VOLATILE MEMORY SELECTION

SEAKR Engineering generally builds Non-Volatile memory cards using commercial FLASH memory. However, considering the requirements of the X2000 program, several other types of non-volatile memory were considered. A brief description of the non-volatile memory technologies considered are listed below.

FRAM (Ferroelectric RAM) uses the ferroelectric effect for a storage mechanism. A FRAM memory cell is made by depositing a film of ferroelectric material in crystal form between two electrode plates to form a capacitor. Data is stored within the crystalline structure. FRAMs have a design similar to other RAM devices in that it

reads and writes simply and easily, but differs as no applied power is required to hold the data state. FRAM data is read using a switched charge that determines the state of the ferroelectric memory. In the process of reading the data, the data is changed; therefore, additional circuitry must be used to re-write the original data back to the memory cell. FRAM uses a direct 5V supply. FRAM generally shows high tolerance to both Single Event and Total Dose Radiation effects. FRAM is capable of a high number (~10 billion) read/write cycles. FRAM is an emerging memory technology and is available from a few sources as low-density (256 Kbit) parts

GMRAM, or Giant Magnetoresistive RAM, is an emerging technology that uses magnetic spin to record data bits. GMRAM has the potential for densities similar to SRAM with comparable read/write cycle times. For all it's promise, GMRAM still has some technical difficulties to work through; reducing drive currents, eliminating cell instabilities due to magnetization vortices, and improving operations at nanometer densities. The biggest problem for use on the X2000 data recorder is availability of parts. It will be several years before GMRAM is available in economic quantities.

FLASH Memory uses floating gate technology for the non-volatile memory cell. Reading and writing to FLASH requires more complicated programming than DRAM, resulting in slower performance. FLASH usually contains a low number of bad blocks when manufactured and can develop additional bad blocks over the lifetime of the part. FLASH is susceptible to radiation effects and can generate electro-magnetic noise that can interfere with other spacecraft instruments. FLASH memory is widely available from several commercial sources in densities up to 256 Mbit.

After an evaluation of these devices, FLASH memory was selected for use in the X2000 NVMS memory board. This decision was based mainly on part density and availability.

3. FLASH CHALLENGES

The non-volatile aspect of FLASH memory is accomplished by using floating gate technology. A floating gate cell consists of a FET with an insulated floating gate located between the control gate and the substrate. An embedded state machine controls each cell. One of the more interesting things about FLASH memory is that the cells can only be changed from 0 to 1 state by an erase operation and can only be changed from a 1 to a 0 state by a write operation. Therefore, reprogramming usually requires an erase cycle followed by a programming cycle. This requirement results in a slower read/write cycle than that required for an equivalent capacity DRAM.

Radiation capability of FLASH Memory is mixed in regard to Single Event Upset (SEU), Single Event Latchup (SEL), and Total Ionizing Dose (TID). FLASH devices are relatively immune to SEU as the devices are only susceptible to SEU when the device is undergoing a read/write/erase cycle. Since FLASH is not powered during most of it's life, the opportunity for a SEU event is very small. Destructive SEL is a concern. The susceptibility of FLASH memory to SEL is highly dependent on the component's design and the manufacture's processes. SEAKR has seen wide differences between different manufacturers in regard to SEL. TID is also a concern for FLASH memory. In continually biased FLASH memory systems total TID tolerance can be less than 5 Krads. If the FLASH is unbiased when not in use, the TID tolerance is greatly improved, into the 35 to 50 Krad range[1].

Commercial FLASH memory is supplied with bad blocks. Rather than screen out devices with defective memory blocks, FLASH memory manufacturers have elected to sell devices with a limited number of bad blocks. This reduces the need for redundant circuitry to ensure the entire memory can be utilized within a device. The location of each bad block is provided by information written to each part. It is considered more cost effective to manage the bad memory locations than produce parts that have 100% good memory. Commercial specifications typically permit up to 20 bad blocks in a 2048 block device. While this philosophy works well in the commercial world, it leads to unnecessary memory waste in spaceflight systems.

FLASH devices require a high current during write and program cycles. This current generates a significant electromagnetic field that can inject noise into the system. For most applications this noise goes unnoticed, but in space probes this EMI can affect nearby sensitive scientific instruments.

The number of Write/Erase cycles is limited for FLASH memory. This is due to gate oxide damage that can accumulate during each cycle. The limit is often improved as the manufacturer's process matures. In mature products, the published limit on write/erase cycles may be due to the inability to accumulate enough test cycles to prove the actual physical limit.

4. FLASH TRADE-OFFS ON X2000

The NVMS design had to account for each of the challenges presented by the selection of FLASH devices as our storage media.

Writing to multiple FLASH devices (10) simultaneously minimized the speed limitations. In addition, the X2000 NVMS design includes two banks of memory, each with a dedicated 2-Kbyte First-In First-Out (FIFO) buffer, that

can be alternately written allowing both banks to be programmed simultaneously.

Radiation concerns were satisfied by a series of SEL and TID testing on devices from several manufactures. This testing found the Samsung 128 Mbit FLASH provide the greatest imunity to SEL and TID[1]. The TID radiation performance limitation was handled by enclosing the FLASH memory devices inside a 0.25 inch Tungsten shell. This shell increases the TID tolerance by approximately a factor of four (4). In addition, the NVMS design includes a power switching capability for the FLASH devices allowing operational planning to further increase the FLASH TID tolerance by powering down the devices when not in use.

To reduce the number of bad blocks in spaceflight the NVMS, SEAKR selects parts using tighter criteria than commercial processes. FLASH memories for space systems include additional screening of the devices to "cherry pick" parts that have fewer bad blocks. This process reduces the lot yield but increases available memory. Past experience has show low yield loss for devices with 15 or less blocks. The X2000 NVMS maximized available memory by screening devices to 10 blocks or less.

SEAKR selects FLASH devices only from mature product lines minimizing read/write cycle limitations. The effects of wear-out are further mitigated by software leveling routines. These routines control the wear on a given block by distributing the data evenly throughout the device. Thus exercising the entire device and not just one area. Though effective in wearing the device evenly, it is still subject to the fundamental limited life problem intrinsic in the FLASH technology.

Careful consideration must be given to power and ground plane noise immunity. FLASH devices exhibit a fairly high instantaneous current during write and program cycles, generating considerable noise in the system. This problem can be reduced by PCB layer placement, routing techniques, and decoupling capacitors. Power and ground planes are used in the NVMS. Loop area is kept to a minimum. Large and small bypass capacitors are used and placed very close to power and ground pins of the devices. This offers low impedance to ground in different frequency domains as well as charge storage. PCB layer placements are such that every signal layer has a reference plane.

5. CONCLUSIONS

A well-designed non-volatile memory system based on FLASH memory can meet the requirements of interplanetary science missions. Design of this system must compensate for FLASH's shortcomings in speed, radiation tolerance, noise, and read/write cycle life. Compensation for these shortcomings does result in cost to the mission in complexity, weight, performance and dollars.

These costs could be avoided if a non-volatile memory was available that performed similar to standard DRAM. The FRAM and GMRAM considered for use on X2000 continue to hold promise. Newer technologies such as chalcogenide based RAM and battery backed up SRAM are also being considered for future designs. SEAKR will incorporate these new technologies into our PCI NVMS when it becomes technically and economically feasible.

By understanding the advantages and limitations of FLASH memory, SEAKR Engineering is able to construct a PCI Non-Volatile Memory System suitable for interplanetary missions using currently available technology.

---

[1] D. R. Roth, J. D. Kinnison, et al, "SEU and TID Testing of the Samsung 128 and the Toshiba 256 Mbit FLASH Memory", 2000 IEEE Nuclear and Space Radiation Effects Conference.

Expectations of MRAM in comparison with other non-volatile memory technologies

K.-M.H. Lenssen

Philips Research Laboratories, Prof. Holstlaan 4, NL-5656 AA Eindhoven, The Netherlands,

+31-40-2742754, kars-michiel.lenssen@philips.com

G.J.M. Dormans

Philips Semiconductors, Gerstweg 2, NL-6534 AE Nijmegen, The Netherlands,

+31-24-3533376, do.dormans@philips.com

R. Cuppens

Philips Research Laboratories, Prof. Holstlaan 4, NL-5656 AA Eindhoven, The Netherlands,

+31-40-2743750, roger.cuppens@philips.com

Abstract— Magnetic Random Access Memory (MRAM) is often presented as the ideal, all-purpose solid-state memory of the future. Indeed, it is expected that MRAM may provide a combination of properties that at present are only found distributed over different types of memories (SRAM, EEPROM, Flash, DRAM, etc.). This is certainly a very attractive perspective; if two (or more) different memory technologies could be replaced by a single memory technology (c.q. MRAM), the development costs for a next generation of a product comprising embedded memory could decrease dramatically. However, for this to come true, it is important that the specifications of MRAM will not merely be a compromise, only suitable for certain niche markets, but that they will really be able to compete with the (future) state-of-the-art "conventional memories". In order to obtain a better insight in this matter, we have looked at (expected) properties of different non-volatile memories for one and the same technology node (0.18 μm). From this, the weaknesses and strengths of MRAM have been identified. Special attention was paid to a comparison with ferro-electric RAM (FERAM), which seems to be the most competitive alternative memory technology.

TABLE OF CONTENTS

1. INTRODUCTION

At present the market for non-volatile memories is booming, a.o. because of their use in portable applications. Flash is still the leading non-volatile memory technology (in 1999 Flash memory increased by 83% and totaled $4.6 billion [1]). However, one can wonder for how long this can continue, since there are definitely at least some inconveniences connected to Flash technology, like the power consumption, the limited endurance, the complications with low-voltage applications and the rapidly increasing number of masks that are required to add embedded Flash to a CMOS wafer. Although predictions of the reach of fundamental limits of Si technologies have been around for decades and sofar always have turned out to be false, there is now a big interest in a possible successor of Flash memory, in particular for technology nodes beyond the 100 nm one. On the other hand, Magnetic Random Access Memory (MRAM) technology is getting a lot of publicity, being presented as the ideal, all-purpose solid-state memory of the future. Therefore, it is useful to make a critical survey of the benefits and drawbacks of MRAM technology, which is one of the promising, emerging candidates.

2. MAGNETIC RAM PRINCIPLES

An MRAM is a memory (1) in which the stored data are represented by magnetization directions and (2) in which the read-out is done by a resistance measurement [2]. This definition implies that its operation relies on a magnetoresistance phenomenon. Early MRAMs were based on the anisotropic magnetoresistance (AMR) effect.

Since the amplitude of the AMR effect in thin films is typically less than 5%, the use of AMR-based MRAMs was limited mainly to military and space applications.

The discovery in 1988 of a larger magnetoresistance effect [3], therefore baptized giant magnetoresistance (GMR), changed this situation. The GMR effect allowed the realization of smaller elements with a higher resistance and a larger MR effect (5 à 15%), and therefore a higher output signal. This enabled, in principle, the realization of MRAMs for general applications. A decade after its discovery the GMR effect is already applied in commercial products like HDD read heads and magnetic sensors [4].

A breakthrough in the field of magnetic tunnel junctions around 1995 [5] improved the perspectives of MRAM even further, when a large tunnel magnetoresistance (TMR) effect was demonstrated at room temperature. Since then TMR effects with amplitudes up to >50% have been shown, but because of the strong bias-voltage dependence, the useable resistance change in practical applications is at present around 25%.

In general, both GMR and TMR result in a low resistance if the magnetization directions in the multilayer are parallel, and in a high resistance when the magnetizations are oriented antiparallel. In TMR multilayers the sense current has to be applied perpendicular to the layer planes (CPP) because the electrons have to tunnel through the barrier layer; in GMR devices the sense current usually flows in the plane of the layers (CIP), although a CPP configuration might provide a larger MR effect, since the resistance perpendicular to the planes of these all-metallic multilayers is very small. Nevertheless, supported by the rapidly continuing miniaturization, the possibility to base MRAMs on CPP GMR recently got serious attention [6].

For the representation of the bits there exist different possibilities. So-called pseudo-spin valves [7,8] comprise two ferromagnetic layers that switch their magnetization direction at different magnetic fields; this can be accomplished by using layers of different magnetic materials, or layers of the same material but of different thickness.

In another category of materials, a.o. (exchange-biased) spin valves the magnetization direction of one of the magnetic layers is so rigid that it can be considered fixed under normal operation conditions (this can, for example, be achieved by using exchange biasing [9] or an artificial antiferromagnet [10]).

The addressing of the MRAM is done by means of an array of crossing lines. Writing a certain cell is equivalent to setting a magnetization in the desired direction (for example, magnetization to the left means '0' and magnetization to the right means '1'). By applying a current pulse to a bit line and a word line a magnetic field pulse is induced. Only the MRAM cell at the crossing point of both lines experiences the maximum magnetic field (i.e. the vectorial addition of the fields induced by both current pulses) and its magnetization is reversed; all other MRAM cells below the bit or word line are exposed to the significant lower field that is caused by a single current pulse and will therefore not change their magnetization directions.

The read-out method depends on the type of MRAM. In the case of pseudo-spin valves a number of cells (N) can be connected in series in the word line, because the resistance of these completely metallic cells is relatively low. This provides the interesting advantage that only $1/N$ switching element (usually a transistor) is needed per cell. The associated disadvantage is that the relative resistance change is divided by N. The read-out is done by measuring the resistance of a word line (with the series of cells), while subsequently a small positive plus negative current pulse is applied to the desired bit line. The accompanying magnetic field pulses are between the switching fields of the two ferromagnetic layers; thus the layer with the higher switching field (the data-storing layer) will remain unchanged, while the magnetization of the other layer will be set in a defined direction and then be reversed. From the sign of the resulting resistance change in the word line it can be seen whether a '0' or a '1' is stored in the cell at the crossing point the word and the bit line.

In the case of spin valves with a fixed magnetization direction the data are stored in the other, free magnetic layer, which of course should not be disturbed by the read-out. In this case the absolute resistance of the cell is measured (if desired, differentially with respect to a reference cell). This cell is selected by means of a switching element (usually a transistor), which implies that in this case one transistor is required per cell.

Like other memory technologies MRAM also offers in principle the possibility to store multiple bits per cell [11].

3. PRESENT STATUS OF MRAM

At present, to our knowledge, Honeywell is the only company that produces and sells MRAM; these are based on AMR and GMR pseudo-spin valves. A 1 Mb memory has been announced [12]. Many other companies (a.o. Motorola, IBM, Hewlett-Packard, Infineon, Toshiba) are actively developing MRAM technology, focussing mainly on TMR. Motorola [13,14] and IBM [15,9] have demonstrated arrays of TMR MRAMs with a size of respectively 512 bits and 1 kb. Motorola and Hewlett-Packard have recently announced MRAM production within 3 to 5 years, while USTC has even announced production already in the first quarter of 2001.

Besides in industrial laboratories there is also a lot of research on MRAM at universities and research institutes [16].

0.18μm rules 16Mb low Vdd	EEPROM	Flash			FERAM		MRAM	
		1T CHE/FN	2T	NAND 16 cells	1T1C	2T2C	GMR	TMR
Performance								
Rand Read Access	50-100ns	50-150ns	50-100ns	10-20μs	<100ns	<100ns	<100ns	<100ns
Burst Read Access		10-20ns	10-20ns	10-20ns	<100ns	<100ns	<100ns	<100ns
W/E Time	2ms/Byte 2ms/Byte	10μs/Byte 1s/sect	5ms/page 0.1s/sect	5ms/page 0.1s/sect	<100ns	<100ns	<100ns	<100ns
W/E Voltage	15/15 V	10/15 V	15/15 V	15/15 V	<1-Vdd	<1-Vdd	>1-Vdd?	>1-Vdd?
W/E Current	nA/nA	100μA/nA	nA/nA	nA/nA	nA	nA	mA	mA
Average power	medium	high	medium	medium	low	low	medium	medium
Low Vdd read	+	-	+	-	+	+	+	+
Endurance	1E6	1E5	1E5	1E5	1E12	1E12	unlimited	unlimited
Reliability	+	+	+	+	+/-	+	++	++
Cost aspects								
Cell size	2-3 μm²	0.4 μm²	0.78 μm²	0.37 μm²	0.8 μm²	2-3 μm²	1.3 μm²	1.5 μm²
Extra masks	9	8 – 10	9	9	3	3	3	+/-
Periphery overhead	+	-	+	-	-	+	-	
Cost of Ownership	+	+	+	+	- -	- -	-	
General								
CMOS Compatible	+	+	+	+	- -	-	-	- -
Maturity	+ +	+ +	+	++	-	+	++?	+?
Scalability	-	-	+/-	+/-	++	+	+/-?	+/-?
Testing	+	-	+	+	+?	+?		
Functionality	NV data	NV code	NV code	NV content	NV DRAM SRAM	NV DRAM SRAM	NV DRAM SRAM	NV DRAM SRAM

Comparison of the (expected) characteristics of different types of embedded nonvolatile memories (EEPROM, Flash, FERAM and MRAM) for the design rules of 0.18 μm CMOS technology (T=transistor, C=capacitor, CHE=Channel Hot Electron, FN=Fowler-Nordheim, W/E=write/erase)

4. COMPARISON WITH OTHER NON-VOLATILE MEMORY TYPES

At the moment very positive and promising publications have appeared about MRAM, and the subject has even reached the popular press. Nevertheless, it is still difficult to get an objective and realistic impression of MRAM, since the properties depend on the type of MRAM and it is unclear which of the mentioned advantages can be obtained in one and the same device and how these compare with other memory types. Moreover, it is important to compare memories at a certain, same moment in time; it is not fair to compare future MRAMs with present Flash memories. Therefore we decided to make an inventory of (expected) specifications of the most important non-volatile memory types, in order to get an insight in the strengths and weaknesses of MRAM. We compared MRAM with EEPROM, Flash and ferroelectric RAM (FERAM), for several architectures differing in the numbers of transistors (T) and capacitors (C) per cell; the 2T-per-cell technology for Flash is Philips' choice for low-power and low-voltage embedded Flash memory [17]. In this comparison we aimed at an embedded memory obeying the rules of 0.18 μm CMOS technology.

Our results are summarized in the table at the top of this page and will be discussed in the next section.

5. EXPECTED STRENGTHS AND WEAKNESSES

Of course, the exactness of all numbers in our table can be discussed, but we believe that the data indicate the relative strengths and weaknesses of the considered non-volatile memory technologies well.
From our study the following strengths of MRAM show up:
- *Read and write speed*
  MRAM promises to be very fast. In the table we stated that MRAM will be faster than 100 ns. The actual time in which a magnetization direction can switch is much shorter (<1 ns), so the basic effect is not the limiting factor. However, (like for e.g. FERAM) the time for addressing, read-out etc. will be dominant and since this depends on array size, interconnects, lead resistances etc., we decided not to fill in a concrete number but state that it will be much faster than Flash. The recent demonstrations of Motorola and IBM, which showed read/write times

of less than 10 ns in mini-arrays, support this expectation. In particular, erasing a bit in MRAM can be many orders of magnitude faster than in Flash or EEPROM.

- *Low voltage*
  MRAM does not have the high-voltage demands of Flash and EEPROM.
- *Excellent endurance*
  MRAM promises the best endurance (with respect to number of write cycles) of all memory types, since till now no deterioration mechanism is known.
- *Low number of extra masks*
  One of the most attractive advantages is the fact that only 3 or 4 extra mask steps are needed to add embedded MRAM to a CMOS wafer. For Flash memory this number is about two to three times higher and, moreover, seems to be increasing with every technology generation. This also suggests that MRAM will be a low-cost technology (assuming that the required silicon area is similar).
- *"All-purpose" memory*
  The most intriguing and spectacular promise of MRAM is to replace more than one of the present memory types. The expected properties are such that it can replace non-volatile memories like Flash and EEPROM, but besides that, MRAM may also allow to substitute SRAM and DRAM. The speed of MRAM is such that it can compete with (most) SRAMs, while MRAM can be much smaller and therefore more inexpensive and is non-volatile. On the other hand, several companies have claimed that the cell size of MRAM can compete with DRAM [8,9], while offering the advantage of non-volatility. So on top of offering specific advantages over certain memory types, it promises a big cost reduction for memory-containing products like smartcards by limiting the design and development efforts to one instead of several type of memories.
- *Good miniaturization possibilities*
  There seem to be possibilities to scale MRAM significantly beyond the 100 nm technology node. A limitation will probably be formed by thermal relaxation.

These promises of an inexpensive, fast, dense, embedded non-volatile memory is obviously very attractive. However, we also identified the following weaknesses or possible issues of MRAM:

- *Cell size*
  By adhering strictly to the design rules in 0.18 μm technology for the upper layers, we estimated cell sizes of 1.3 μm^2 and 1.5 μm^2 for respectively GMR and TMR. This is several times larger than some Flash-type memories. Nevertheless, others have claimed that MRAM can be comparably dense or even denser than DRAM [6,8,9]. Although it is virtually impossible to deviate from design rules for embedded memories, we also see possibilities for cell size reduction, since in our estimation the cell size was mainly determined by the metal line width that is

required because of the current pulses. For example, the use of copper metallization might enable smaller cells.

- *Write current*
  Besides its influence on the cell size, the required write current can also pose requirements on the addressing and peripheral electronics. Also it may have a negative effect on the power consumption. Fortunately the current pulses may be only very short and therefore the negative consequences may be limited.
- *Temperature stability*
  CMOS processes usually include an alloy step at or above 400°C. Considering our experiences on the thermal stability of GMR sensors [18] this seems to be a serious issue for embedded memories, if one wants to adhere to the standard CMOS process. However, for MRAMs it is only required that they survive the exposure to high temperatures, while for the sensors normal operation at elevated temperatures was tested.
- *Resistance*
  Scaling of the resistance of TMR materials seems to require an ongoing reduction of the barrier layer thickness. Since this thickness is already in the order of several monolayers, one may wonder whether resistance scaling can keep up with Moore's Law. While GMR does not have this problem, the reliable and fast sensing of the small resistance change is not trivial.
- *Maturity*
  While GMR is already used in commercial products like read heads and sensors, TMR is still less mature.

The abovementioned strengths constitute certainly significant advantages over the common silicon non-volatile memories EEPROM and Flash. However, most of the listed advantages are also expected from FERAM and from the table the differences in merits of FERAM and MRAM are less clear.

Nevertheless, there are some problems or difficulties connected to FERAM that MRAM does not have, like:

- problems with integration with silicon because of the need to anneal the ferroelectric material at high temperature,
- sensitivity of the ferroelectric material to H_2,
- the ferroelectric materials are usually toxic and not environmentally friendly (e.g. because they contain lead),
- the limited endurance; this is likely to be fundamental since ferroelectricity requires movement of an atom, while magnetoresistance only relies on the non-materialistic rotation of spins,
- destructive read-out, which also implies that the read endurance is restricted by the write endurance,
- the scaling to very small dimensions is not obvious since ferroelectricity is a "bulk property" of the material.

29

Of course, MRAM technology poses other difficulties, like the earlier mentioned issues, but these seem less severe than those of FERAM.

6. CONCLUSIONS

We have given a short overview of the principles and present status of MRAM, and presented the results of a comparison of MRAM technology with other non-volatile memories (Flash, EEPROM and FERAM). Although the listed weaknesses of MRAM still need to be overcome, it is clear that MRAM is expected to give revolutionary advantages over the currently used silicon embedded memories. FERAM can also provide some similar advantages, but it shows some problems that MRAM does not have. In particular, the low required number of extra masks and the fact that MRAM may be able to be substituted for "all" types of memories (also e.g. SRAM or DRAM) are very attractive, particularly for embedded memories and systems-on-a-chip.

REFERENCES

[1] M. Villeneuve, "SIA says chip sales surge", Solid State Technology 43(3), 12 (2000)

[2] For a recent overview about MRAM see, for example, M. Johnson, "Magnetoelectronic memories last and last...", IEEE Spectrum, February 2000, 33 (2000) or J. Daughton, "Magnetoresistive Random Access Memory (MRAM)", http://www.nve.com, (2000)

[3] G. Binasch, P. Grünberg, F. Saurenbach and W. Zinn, "Enhanced magnetoresistance in layered magnetic structures with antiferromagnetic interlayer exchange", Phys. Rev. B39, 4828 (1989); M.N. Baibich, J.M. Broto, A. Fert, F. Nguyen Van Dau, F. Petroff, P. Etienne, G. Creuzet, A. Friederich and J. Chazelas, "Giant magnetoresistance of (001)Fe/(001)Cr magnetic superlattices, Phys. Rev. Lett. 61, 2472 (1988)

[4] For a recent overview about GMR sensors see, for example, K.-M.H.Lenssen, D.J. Adelerhof, H.J. Gassen. A.E.T. Kuiper, G.H.J. Somers and J.B.A.D. van Zon, "Robust giant magnetoresistance sensors", Sensors&Actuators A85, 1 (2000)

[5] J.S. Moodera, L.R. Kinder, T.M. Wong and R. Meservey, "Large magnetoresistance at room temperature in ferromagnetic thin film tunnel junctions", Phys. Rev. Lett. 74, 3273 (1995)

[6] J.-G. Zhu, Y. Zheng and G.A. Prinz, "Ultrahigh density vertical magnetoresistive random access memory", J. Appl. Phys. 87, 6668 (2000)

[7] A. Pohm, B. Everitt, R. Beech and J. Daughton, "Bias field and end effects on the switching thresholds of 'pseudo spin valve' memory cells", IEEE Trans. Magn. 33, 3280 (1997); B. Everitt, A. Pohm, R. Beech, A. Fink, and J. Daughton, "Pseudo spin valve MRAM cells with sub-micrometer critical dimensions", IEEE Trans. Magn. 34, 1060 (1998)

[8] S. Tehrani, J.M. Slaughter, E. Chen, M. Durlam, J. Shi and M. DeHerrera, "Progress and outlook for MRAM technology", IEEE Trans. Magn. 35, 2814 (1999)

[9] S.S.P. Parkin, K.P. Roche, M.G. Samant, P.M. Rice, R.B. Beyers, R.E. Scheuerlein, E.J. O'Sullivan, S.L. Brown, J. Bucchigano, D.W. Abraham, Y. Lu, M. Rooks, P.L. Trouilloud, R.A. Wanner and W.J. Gallagher, "Exchange-biased magnetic tunnel junctions and application to nonvolatile magnetic random access memory", J. Appl. Phys. 85, 5828 (1999)

[10] U.K. Klostermann, S. Zoll, H.A.M. van den Berg, G. Rupp, R. Kinder, J. Bangert, L. Bär, R. Leuschner, G. Gieres and J. Wecker, "MRAM: fabrication and magnetic properties of sub-micron tunnel junctions", in Proceedings Statusseminar Magnetoelektronik (14-16 June 2000, Dresden, Germany), 83 (2000)

[11] R. Zhang, M.M. Hassoun, W.C. Black Jr., B. Das and K.A. Wong, "Demonstration of a four state sensing scheme for a single pseudo-spin valve GMR bit", IEEE Trans. Magn. 35, 2829 (1999); W.-C. Jeong, B.-I. Lee and S.-K. Joo, "Three level, six state multilevel magnetoresistive RAM (MRAM)", J. Appl. Phys. 85, 4782 (1999)

[12] T. Zhu, "Development of MRAM for space and commercial applications", presentation DA-03 at the Intermag 2000 (April 9-13, 2000, Toronto, Canada)

[13] M. Durlam, P. Naji, M. DeHerrera, S. Tehrani, G. Kerszykowski and K. Kyler, "Nonvolatile RAM based on magnetic tunnel junction elements", 2000 IEEE Int. Solid-State Circuits Conf., Digest of Technical papers, San Francisco CA, U.S.A., 7-9 Feb. 2000, 130 (2000)

[14] S. Tehrani, E. Chen, M. Durlam, M. DeHerrera, J.M. Slaughter, J. Shi and G. Kerszykowski, "High density submicron magnetoresistive random access memory", J. Appl. Phys. 85, 5822 (1999)

[15] R. Scheuerlein, W. Gallagher, S. Parkin, A. Lee, S. Ray, R. Robertazzi and W. Reohr, "A 10 ns read and write non-volatile memory array using a magnetic tunnel junction and FET switch in each cell", 2000 IEEE Int. Solid-State Circuits Conf., Digest of Technical papers, San Francisco CA, U.S.A., 7-9 Feb. 2000, 128 (2000)

[16] See, for example, H. Boeve, C. Bruynseraede, J. Das, K. Dessein, G. Borghs, J. De Boeck, R.C. Sousa, L.V. Melo and P.P. Freitas, "Technology assessment for the implementation of magnetoresistive elements with semiconductor components in magnetic random access memory (MRAM) architectures", IEEE Trans. Magn. 35, 2820 (1999)

[17] T. Ditewig, R. Cuppens, K.-L. Chen, V. Frowijn, F .Jetten, W. Kalkman, M. Malabry, A. Slenter, M. Storms, N. Tandan, S. Teuben and A. Grácio, "An embedded 1.2-V read Flash memory module in a 0.18 μm logic process", to be published

[18] K.-M.H. Lenssen, A.E.T. Kuiper, J.J. van den Broek, R.A.F. van der Rijt and A. van Loon, "Sensor properties of a robust giant magnetoresistance material system at elevated temperatures", J. Appl. Phys. 87, 6665 (2000)

BIOGRAPHIES

Kars-Michiel Lenssen was born in Utrecht, The Netherlands in 1968. He received the M.Sc. degree in Applied Physics from Eindhoven University of Technology in 1989. In 1994 he received his Ph.D. degree from Delft University of Technology. Since 1994 he has been a senior scientist at the Philips Research Laboratories in Eindhoven, working on the giant magnetoresistance effect and on magnetic devices. After research projects on the subject of (giant) magnetoresistance sensors, he is currently leading an investigation on Magnetic RAM.

Do Dormans (The Netherlands, 1958) received his PhD degree from the Technical University in Eindhoven (The Netherlands) in 1987. In that year he joined the Philips Research Laboratories in Eindhoven, The Netherlands where he worked on CVD, ferroelectric technology and non-volatile memories. Since 1999 he works as senior device engineer non-volatile memories at Philips Semiconductors in Nijmegen, The Netherlands.

Roger Cuppens was born in Zonhoven, Belgium in 1948. He received his electrical engineering degree and his Ph.D. in applied science from the Katholieke Universiteit Leuven, Leuven, Belgium, in 1972 and 1980 respectively. In 1972 he was employed as a Research Assistant at the E.S.A.T. Laboratory of the Katholieke Universiteit Leuven. In 1979 he joined Philips Research Laboratories, Eindhoven, The Netherlands where his main activity was in the design of memories. Currently he is involved, as a Principal Scientist, in the development and design of embedded non-volatile memories.

Opportunities and Challenges for

Embedded Flash Memory

Sohrab Kianian
Silicon Storage Technology, Inc.
Sunnyvale, California 94086
skianian@ssti.com

David Sweetman
P.O. Box 189
Dyer, Nevada 89010
d-dsweetman@att.net

Abstract— With advances in deep submicron CMOS technology, faster timing and more feature-rich integrated silicon devices are being used in consumer electronics advanced communication and networking systems, computers, servers, and virtually all other electronic systems. The demand for performance and functionality is ever increasing and a key component of that demand is the in-system nonvolatile alterability of both code and data. Larger operating and application codes as well as configuration and personalization codes are stored in various IC components, which require field upgrade capability. Various types of data require nonvolatility. With systems becoming increasing portable and smaller in size, various bulky mechanical elements, such as magnetic storage disks, are designed out and silicon programmable elements are being substituted for them. The silicon elements must continue to provide better reliability and an ability to operate in harsher environments than media that have mechanical movement. These trends have presented an opportunity for flash and embedded flash as never seen before. The question is whether embedded flash is ready to serve these demands and keep up with the projected increasing demands well into the future.

TABLE OF CONTENTS

1. EXAMPLES

Examples of Flash in Portable Electronics

| Code Storage Data storage: Names. Numbers. Preferences. Feature options | Code Storage Data Storage | PDA Lap-top Electronic Alarm Digital Camera Personal DVD |
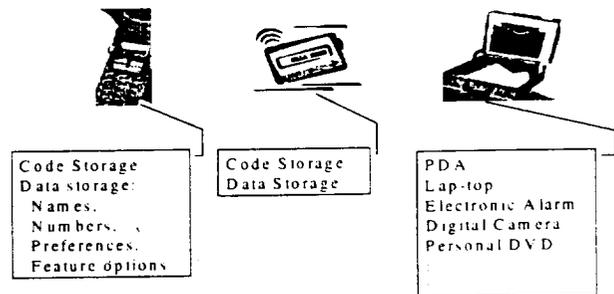
2. EXTRINSIC AND INTRINSIC FACTORS

The process of integrating flash memories into a complex multi-function design, on a monolithic silicon substrate, is driven by the factors of functionality, performance, security, cost, reliability, and portability. Portable, personal communication appliances are typically very small in size and consume very low power in order to operate for long periods of time; and may require other IC device performance factors that sets them apart from competing products in the market place. Such a product, although cost sensitive, is not predominantly designed for its a low cost advantage. While a multiple chip solution may result in lower cost, performance and size dictate an integrated device solution. These factors have different priorities in a desktop product where size and even power may be less of a factor; while device cost, features, and performance are more emphasized.

The technology most suitable for widest range of embedded designs must possess all of the determining factors and be able to produce a distinct advantage in any of the desired applications. While a "one size fit all" solution may seem unachievable or unrealistic, searching for one brings us closer to creating a solution matrix that best incorporates the important factors.

The success of the embedded flash business is driven by both extrinsic or macro factors as well as intrinsic or micro factors. Following table summarizes these categories to first order:

Extrinsic factor	Intrinsic factor
1. Availability of reusable embedded flash blocks from multiple sources with ample wafer supply	1. Power consumption
	2. Cost
	3. Scalability
2. Reasonable assurance of flash technology roadmap in parity with advanced logic technology roadmap	4. Performance & Flexibility
	5. Reliability
3. Integration cost, exclusive of flash process adder	6. Security
4. Hassle-free IP business model	7. Ease of integration & logic compatibility
One-stop shopping at multiple foundries	
Availability for internal second souring	
Proven technology track record	
Comprehensive support model	

While the market for flash memories and semiconductors in general is subject to fluctuations and can cause the extrinsic factors to become more or less dominant, some considerations are always relevant. For instance, depending on the memory architecture, multiple layer metalization may or may not be required. When the added nonvolatile memory does not require as many metalization layers as the base process into which it is integrated, then one has to consider how much memory integration is justified to amortize the added metalization cost on the whole chip. In these cases, the memory density is small. This consideration is over and above the added cost of nonvolatile memory process (intrinsic). Typically more complex IC's (e.g., mixed mode IC, DRAM) have a large process overhead (e.g., more poly or metal layers) compared with a typical base logic, regardless of the nonvolatile memory. In such cases the added nonvolatile memory cost adds to the already expensive process and may only be fractionally important. The yield and ease of use and debug for the nonvolatile memory, then becomes paramount since any yield fall-out on NVM portion magnifies the cost of production.

In essence, for complex IC processes there must be other compelling reasons for integration; such as performance or reliability enhancement, added features, footprint reduction, and power saving.

However, as the overhead cost of the flash is reduced (intrinsic), the cross over point shifts in the more complex process direction. The process overhead cost combined with the silicon utilization (design overhead cost), test and packaging costs determine the cross over point for a given embedded flash technology. Figure 1 illustrates that an economically sound choice can be made given the die size

and the cost of the IC. With a more cost-effective flash solution this choice favors monolithic integration for a much broader range of logic devices.
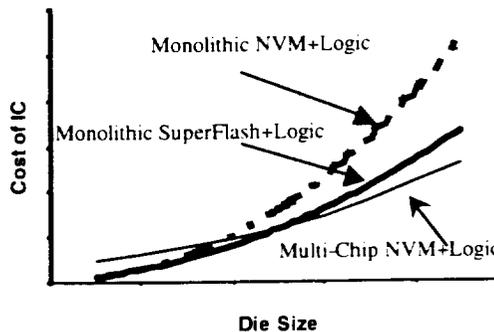

Figure 1: Flash/SRAM Integration Cost Trade-Off

3. OPPORTUNITIES FOR EMBEDDED FLASH

With advent of professional foundries and fabless IC device manufacturers, the landscape of semiconductor IC manufacturing firms, their market positions, and success have become very fluid and more competitive. The non-USA foundries now supply a relatively high percentage of the total semiconductor IC wafers and packages and they are expanding their presence at a relatively rapid paste. In 1999 the wafer production output of the foundries were expected to exceed $5.4 billion dollars, on a worldwide basis. Dataquest is predicting that during this new decade pure play foundries will account for roughly half of all semiconductor production and their revenues will double between 1999 and 2002. Interestingly, Integrated Device Manufacturers (IDMs) as customers represent the fastest-growing portion of the foundry business. IDMs' demand for foundry is expected to triple between 1998 and 2003 to account for roughly one-third of the foundry market. The fabless companies' demand will grow by 18.7% CAGR.[1]

To attract lion's share of this incredibly fast growing segment, foundries are positioning themselves as one-stop shop for processes, design services, IP supplier or dealer, mask making, test, and finish services. The foundries establish Quality Systems per ISO-9001 and perform accelerated stressing on test structures and products in accordance with JEDEC standards. Process control and optimization is rigorous and references various EIA and JEDEC documents. This brings a very attractive value proposition to the foundry customers, fabless and IDM alike. Fabless companies are now free to innovate on the basis of design techniques and unique functionality. At the same time the IDMs are no longer bound by homegrown IP and can leverage their resources in obtaining and using the best-in-class solutions.

33

This paradigm shift in IC industry has created a unique opportunity for embedded NVM memories. New products are being offered at a much faster rate due to flexibility in outsourcing to the foundries. At the same time the need to build programmability into these IC's are increasing as a result. It would be difficult to build new products with mostly the same feature-sets once the code is changed due to system modifications, new applications, etc. without the ability to alter the memory contents. Code storage represents a significant portion of the embedded flash requirement, but the data storage requirements are also very high. Code storage also allows backward compatibility with various items or hardware, so that the functionality of a system can appear the same while the cost of the system is reduced by using more advance (integrated) components. Emulation or configuration of systems occurs in software (stored on-chip), while utilizing the ever-improving components.

4. SUPERFLASH®-TECHNOLOGY OF CHOICE

The above subtitle may seem like a bold claim, but we will try to substantiate that claim with evidence and reason. In both the intrinsic and extrinsic factors, SuperFlash can claim superiority and leadership.

1) Power Consumption:

In reprogrammable nonvolatile memories, battery energy is consumed during memory alteration (Program and Erase) and memory accessing (Read) operations. Therefore, the embedded flash user must examine these operations and related mechanisms in some detail and make appropriate compromises to select the best, i.e., least power consumption, solution for the embedded flash technology. Note power is a function of voltage, current, and time, so all these factors must be considered when calculating power consumption.

a) Low power required for programming:

Source side Channel-Hot-Electron (CHE) injection is »100 times more efficient than Drain-side CHE injection.[2] Therefore, during the program operation (given the same amount of charge is transferred), the power consumed by these types of flash memories (e.g., SuperFlash) is considerably less than those using Drain-side CHE.

b) Low power required for erase:

Power consumed during erase is much lower than stacked-gate because the erase time of SuperFlash is so much faster, i.e., ms not seconds.

c) Verify operations:

Most stacked-gate memories require a convergence algorithm (including some special verify operations) or some form of healing scheme in order to assure the Erase (or Program) thresholds are (and are maintained) within a desired range. Split-gate memories (e.g., SuperFlash) do not require such algorithms or schemes, since a control gate controls a part of the channel. Thus, the associated state machines for these algorithms and schemes are unnecessary for SuperFlash, due to split-gate architecture and resulting immunity to over-erase and hot-hole trapping. Power consumed during the verify operations is important since these verify operations must be fast in order not to degrade the timing performance of the device during program or erase operations.

d) Power consumption during read operation:

Reprogrammable nonvolatile flash memory, embedded in an SOC, is programmed a finite number of times (hundreds to tens of thousands of times are typical) and is constantly read. The Read operation has to be considered a "sustained" event compared to the Erase and Program operations. While power consumption during erasing and programming is a critical consideration, the power use during read is even more critical. The factors determining power use during the read cycle are: 1) cell read current and drive voltage, 2) word-line RC (Resistance-Capacitance) delay, 3) bit-line RC delay, and 4) sense amplifier design. Some circuit design techniques can be universally employed for any memory design to reduce power consumption (e.g., deep sleep mode in conjunction with address transition detection). The parasitic delays of word-line and bit-line can be similar for most types of memories. When the supply voltages continue to scale, achieving an acceptable read timing/power characteristic from the memory can be a challenge. For instance, when the low threshold of the memory transistor needs to be substantially positive, then word-line boosting is employed to read the memory transistor, when the supply voltage and the threshold voltage are not significantly different. This scenario is very much the case for 0.25um and smaller geometry for the stacked-gate type memories. This memory cell type need a positive (low) threshold in order to allow programming of the selected cells sharing the same bit-line with the low-threshold unselected cells. Split-gate devices, in particular SuperFlash, operate with negative low thresholds. Word-line boosting is only necessary for the split-gate memory to overcome the natural control gate threshold. This can only become necessary in sub 0.13um geometry generations. In essence, for the same supply voltage (in sub 0.25um generations) the cell read power consumption characteristics are more favorable for

34

a split-gate device compared to a stacked-gate device. This advantage also allows read operation across a wider temperature range, e.g., for a harsher environment.

2) Cost

The reprogrammable nonvolatile memory added cost to the CMOS process is determined by several factors. Complexities are added by the introduction of new layers (e.g. floating gate, high voltage transistors gate oxide, and high voltage node junction implants). In addition to these added layers, added circuits, testing, debugging, and yield enhancement costs are required specific to the NVM.

SuperFlash offers multiple functionality (optimized cost versus functionality trade-off) in one process: it can be designed as an OTP, Flash, or E2PROM with only supporting circuitry modified to achieve desired functionality. Furthermore, SuperFlash can well integrate with existing base CMOS process without significant alteration, leaving the logic simulation models largely intact. Added layers are the minimum required and standard processes can be used. Operational simplicity allows for simple peripheral circuit requirements, which then enhance yield and debug process. This point bears emphasizing, since SuperFlash does not need a state machine or algorithmic operations for erase or programming; thus, the peripheral logic is greatly simplified.

The split-gate (SuperFlash) cell design lends to a flexible architecture requiring a few simple peripheral circuits to support the memory operations. Fundamentally the technology is based on thick oxides for charge transfer, which assures future scaling capability. The coupling (floating) gate oxide is never subject to a high electric field stress (strong coupling to high voltage node); thus, remains reliable during the life of the device. For a given photolithography capability, the memory cell is longer (compared to stacked-gate) in the bit-line direction due to presence of the split gate. However, the word-line direction is shorter than stacked-gates, since there is no high voltage bit-line isolation requirement (high voltages are applied in the row direction only), nor is there a requirement to achieve high coupling from the word-line.[2] Overall the cell size is comparable to that of the stacked–gate for a given photolithography, but the higher array efficiency due to a simpler periphery, results in smaller die area to achieve the same density and functionality. The R&D cycle time and cost is reduced because the simpler design results in easier debug, reduced test, and faster yield enhancement efforts.

a) Integration Trade-Off:

In the beginning we discussed the integration trade-off, especially as they relate to the cost of running NVM

not required metalization steps over larger NVM chip areas. SuperFlash can be
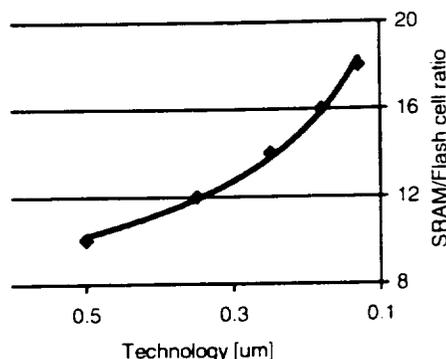
Figure 2: relative SRAM size trend

integrated with logic, SRAM, or mixed mode design rather easily and cost effectively (see Figures 1 & 2 above). However, the IC designer still needs to determine threshold requirements for monolithic versus multi-chip package (MCP). One good figure of merit is to estimate die cost versus packaging and test cost. As a rule of thumb, if the former is less expensive than the latter, then integration makes economic sense.

Another consideration, particularly with respect to monolithic integration of SRAM and flash, is that the SRAM silicon area cost is significantly higher than flash and the trend is continuing (see Figure 2). A $10mm^2$ embedded SuperFlash is generally very cost effective, but a total die area exceeding $80mm^2$ becomes expensive to manufacture.

For logic and flash integration, again the maximum die size, optimum flash size, and the die versus package and test benchmarks can be used to estimate the cost-effective flash integration.

3) Scaling

In stack-gate flash cells floating gate oxides are thinned (as a natural requirement of scaling), yet still must have strong electric field across them. These strong electric fields result in compromising reliability with further scaling. Alternate insulating materials cost more to develop and complicate CMOS process compatibility even more. In thin-oxide memories, the high voltages can scale, so long as the planar oxide is allowed to scale without compromising charge retention. F-N tunneling is a well-established physical phenomenon with predictable exponential (voltage dependent) behavior. A certain minimum oxide field would be required to move charge across a planar oxide in a given time. Scaling the thin oxide beyond ~8nm, without compromising reliability is a technological challenge. In fact, the stacked-gate floating gate oxide has not scaled since near 1um generation.[3] Some are beginning to advocate a less than robust

35

reliability specification to allow further thinning of the oxides. Once the boundary of nonvolatile memory and volatile memory is blurred, no one can predict where *this blurring will lead.*

SuperFlash scaling can be generally summarized as using a constant field, constant coupling ratio approach. The constant field approach assures that the voltages used are proportionally lower for each generation, to allow memory oxides to remain reliable and to allow generation and distribution of Program and Erase high voltages on chip. The constant field scaling approach used in SuperFlash allows for acceptable variation of oxide field from one process generation to another, and across the process variation within a given technology generation. The constant coupling ratio scaling approach is adopted to assure that Program and Erase endurance of the SuperFlash is maintained with each generation. To illustrate, the following table lists the critical scaling features of several generations:

	0.7UM	0.5UM	0.35UM
Coupling Oxide	120A	100A	80A
Tunnel Oxide	350A	300A	250A
Vpp (Program)	12.5v	11v	9v
Vee (Erase)	15v	13.5v	11v
Vdd – max	5.5v	5.5v	3.6v
Vdd – min	2.7v	2.7v	1.8v
Cell Size [μm²]	5.96	3.12	1.7
Word-line to Floating Gate Coupling (CR)	21%	23%	24%
Estimated Peak Coupling Oxide Field [MV/Cm]			
Erase	5.5	5.5	5.0
Program	4.7	5.5	6.0
Read	2.5	3.0	3.5
Estimated Average Tunnel Oxide Field [MV/Cm]			
Erase	2.4	2.7	2.8
Program	1.6	1.2	1.2
Read	1.1	1.3	1.2

TABLE (1) – SuperFlash Cell Scaling

4) Performance and Flexibility

In a typical embedded NVM application, the performance (speed and power) of the NVM block is improved compared to the stand-alone equivalent density. This improvement is partially due to eliminating the pad and I/O structures and the more direct (less delayed) interface with the surrounding circuitry. Furthermore, in benchmarking NVM stand-alone products, the standard tests call for a specified capacitive loading on the pins (which may not be representative of the actual application), while the embedded NVM experiences a well known and consistent loading (substantially less than external test loading), based on the integrated chip design.

With a NVM that has fewer peripheral logic circuits, internal capacitive loading is less and performance is enhanced. The more circuits a signal has to pass through, the more performance degradations are likely to occur. This speed performance improvement is a SuperFlash unique advantage for embedded solutions. The speed performance gains are consistent across a wider operating temperature range.

NVM architecture also plays an important role in determining the performance of the product. For instance, NAND stacked-gate multi-transistor architecture is well known to have inferior read access performance compared to NOR architecture, because a read has to be serially accessed through multiple pass gates. The overall size of the bit-line and word line affects the RC delay; thus, the more flexible the options are for the NVM solution, the better one can optimize the performance of the integrated product.

When designing embedded NVM IC's, designers are concerned about the program and erase performance, as well as the read speed. The fixed program and erase pulse (as utilized in SuperFlash) that is constant through out the life of the IC is more embedded design friendly compared to the algorithmic approach where program and erase times increase over the life of the application.

Single byte or word altering capability (E^2PROM) is necessary in many integrated IC's for configuration, parameter, or user data. With some flash technologies using FN (Fowler-Nordheim) tunneling for program and erase (e.g., DINOR, NAND, AND) this feature is difficult to obtain for reasonable write times since the programming charge transfer mechanism (i.e., FN tunneling) is a slow mechanism. In the split-gate SuperFlash technology E^2PROM capability is readily available through implementing traditional circuit techniques for isolating and selecting single bytes or words.

With technologies utilizing drain-side CHE injection for programming, large programming currents are required; thus, prohibiting multiple-byte parallel programming to effectively reduce program time. In applications where single-byte-altering capabilities are not required (e.g., program code, message data, or look-up tables), source-side CHE injection cells (e.g. SuperFlash) or Fowler-Nordheim tunneling cells (e.g., DINOR, NAND, AND) allow the embedded memory designer to take advantage of parallel programming improve write speed performance.

The ability to integrate flexible sector sizes for variety of applications, in a single design or different

36

designs, provides long-term benefits for the integrator. The designer needs to only to learn one NVM technology and freely modify the design to match the application space. Flexible sector sizes provide a shorter time to market in at least three ways: first by leveraging the previously collected experience to improve future products, second by designing reusable blocks for different application segments, and third by creating test operation uniformity for different devices.

5) Reliability

Reliability of NVM is the ability of the memory to meet the data sheet for the expected life. Above and beyond normal CMOS and package reliability, NVM reliability is measured by 1) ability of the memory to endure specified number of data alterations (endurance), and 2) ability to retain data (with or without power applied) for a specified lifetime (retention).

In stacked-gate technologies the insulators are thin and subject to sustained electrical stress since F-N tunneling through thin, planar oxide is used to remove the floating gate charge. In some stacked-gate implementations (e.g., NAND, AND, DINOR) the average electrical field across the floating gate oxide can exceed 10MV/Cm. Even when positive high voltage is reduced and negative voltage is used to achieve cell operation, the oxide field magnitudes are not reduced and same sensitivity persists. Prolonged applications can cause oxide damage; thus, loss of the charge defining the data states.

Inter-poly tunneling used in SuperFlash cell utilizes field enhancement features that allow tunneling through thick oxide at low average electric field. Furthermore, in this cell the gate oxide under floating gate (i.e., coupling oxide) is not exposed to tunneling oxide electric stress.
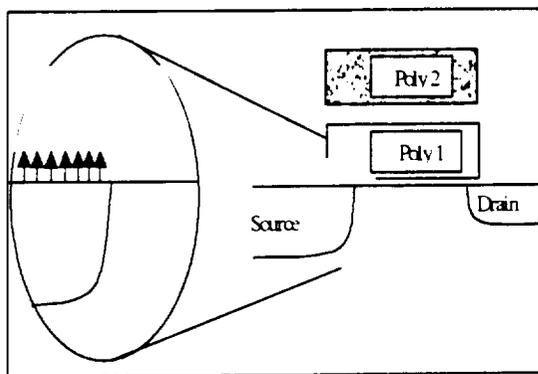
FIGURE 3– Electric Field During Erase (Stacked-gate Flash)

In SuperFlash, the electric field distribution in the thick tunnel oxide is non-uniform [See Figure 4]. In comparison the thin oxide technologies have uniform, strong field through out the tunnel (floating) gate oxide [See Figure 3]. Since the electron trap generation rate is proportional to the electric field, the interpoly dielectric of SuperFlash experiences charge trap site generation only near the "tip" region and the rest of the oxide remains trap free. In stacked gate or thin oxide flash, the generated traps are distributed evenly through out the oxide. These traps can allow the stored charge in floating gate to hop from site to site, depleting the charge stored on the floating gate. In the case of asymmetrical traps, such hopping paths do not exist.
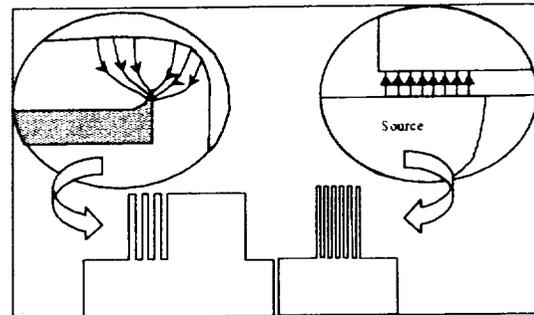
FIGURE 4 – Charge Retention & Trapping (Inter-poly tunneling vs. Stacked-gate)

Endurance is the ability of the memory cell to sustain repeated erase and program and still meet all data sheet operations, including data retention. In thin-oxide memories (i.e., tunneling through floating gate oxide) endurance is limited by oxide leakage or breakdown resulting in charge loss, creation of trapped holes resulting in erratic erase behavior of the memory cell, or the inability of the device to maintain tight erase threshold distribution of the cycled bits. In comparison, the SuperFlash cell endurance limitation is encountered when the trapped charges in the erase oxide cause a gradual increase (the "trap-up") of the required voltage or time to erase the cell effectively. This phenomenon is distinguishable from the thin oxide limitations in several ways. First, the trap-up behavior is not suddenly catastrophic; thus, depending on the application constraints one can increase the number of endurance cycles by either increasing the erase time or voltage. Second, trap-up is a predictable behavior; therefore, a substantial margin can be designed for the erase voltage or time to achieve the desired endurance. Third, as the technology progresses and integrity and chemistry of silicon and oxide processing improve, the trap-up behavior improves.

Far more important than simply looking at the endurance alone, is the ability to retain the required charge after subjecting the memory to numerous cycles of

37

erase and program. Therefore, data retention verification
(i.e., reliability stressing) must occur after the data sheet
specified number of cycles, e.g., 10,000 or more.

Endurance and data retention are validated using the
applicable JEDEC standard test methods and criteria.

4.   CONCLUSIONS

We have broadly reviewed the trade-off and
opportunities for embedded flash memories. The
challenges for wholesale acceptance and propagation of
this technology remain, i.e., availability through multiple
suppliers, aggressive roadmap, repeatability, and being
able to leverage learning and apply that to new designs.
SuperFlash has addressed these issues most effectively, as
SuperFlash is offered through multiple foundry suppliers
and IDMs at multiple technology generations. The state
of the art foundries are being driven by IC designers to
offer the state of the art embedded flash solution and these
forces drive the aggressive roadmap of SuperFlash.
SuperFlash has been in use by many third party designers
for a number of years and has successfully been
demonstrated in multitude of varied products. The
collective knowledge about this unique technology is
amassing and driving a de-facto standardization of it,
because ultimately the primary beneficiaries are the
technology users.

6. REFERENCES

1:        EE Times, Nov. 15, 1999

2:        1994 Symposium on VLSI Technology, pp71-72

3:        1999 IEDM, Session 11, Article 1

4:        "Reliability of Reprogrammable Nonvolatile
          Memory"; Nonvolatile Memory Technology
          Review, 1998

Radiation Issues and Applications of Floating Gate Memories

L. Z. Scheick, 4800 Oak Grove, Pasadena, Ca 91202, leif.scheick@jpl.nasa.gov

D. N. Nguyen, 4800 Oak Grove, Pasadena, Ca 91202, duc.nguyen@jpl.nasa.gov

Abstract—The radiation effects that affect various systems that comprise floating gate memories are presented. The wear-out degradation results of unirradiated flash memories are compared to irradiated flash memories. The procedure analyzes the failure to write and erase caused by wear-out and degradation of internal charge pump circuits. A method is described for characterizing the radiation effects of the floating gate itself. The rate dependence, stopping power dependence, SEU susceptibility and applications of floating gate in radiation environment are presented. The ramifications for dosimetry and cell failure are discussed as well as for the long term use aspects of non-volatile memories.

TABLE OF CONTENTS

1. INTRODUCTION

With a large number of long term space missions being flown with non-volatile memory (NVM) banks as integral avionics systems, such as the upcoming mission to Europa, the need to precisely understand the effects of radiation on the NVM types has grown considerably. This is mostly because the NVM banks on such missions are by far the most susceptible to total ionizing dose (TID) effects, and fairly susceptible to single event effects (SEE) [1, 2]. The variation of radiation effects with variables like dose rate, stopping power , and radiation type has become very important.

Flash memories command the lion's share of commercial NVM applications: digital cameras, wireless communication devices, and computer storage. They are also being considered for many current and future JPL space systems, including solid-state recorders for the X2000 project. Previous missions used DRAMS for solid-state recorder applications. Flash technology has evolved very rapidly during the last five years. Multi-level 128-Mb devices that can store more than one bit per cell, are now available commercially including advanced flash memories with complex internal control circuitry, block erasing and writing to internal buffers, which is all transparent to the user. Flash memories of older generations used an external erase/write pin that required a separate power supply of 12 volts for erasing and writing data to the cells. Newer, single power supply advanced flash memories rely on internal charge pump circuits to provide the high voltages that are needed for erase and write operations. Scaling issues for newer flash devices are more complex than conventional CMOS devices because of the need for high voltage (10 to 20 volts) for erasing and writing. The tunnel oxides used in flash storage cells are above 9nm and have not been reduced, as devices have been scaled [1, 2]. Multi-level cell storage is one approach for the achievement of higher storage density. Ionizing radiation has always been an issue for flash devices, and newer technologies have increased susceptibilities due to scaling.

Two separate technologies have been developed to build the basic cell structure of advanced flash memories. The NOR structure technology uses channel hot electron injection to program and Fowler-Nordheim (F-N) tunneling to erase. A typical floating gate cell is shown in Fig. 1. When a flash memory is read, 3.3V is applied to the control gate and drain while the source is grounded. In an erased cell, the control gate voltage overcomes the transistor turn-on threshold voltage, V_{th}, and the sense-amplifier circuitry detects the drain-to-source current and translates it to a "1". The voltage due to the control gate of a programmed cell is not sufficient to overcome the V_{th} and the absence of the drain-to-source current translates it to a "0".
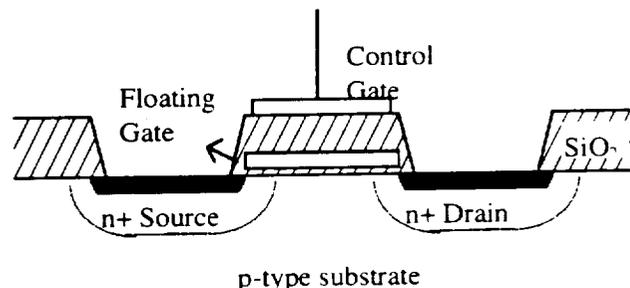
Fig. 1. Standard floating gate device. The drawing is not to scale and the gate oxide thickness varies for each technology.

The NAND structures use F-N tunneling for both writing and erasing. Reading is identical to the NOR structure's method. The NAND structure erases and programs uniformly across the gate oxide, resulting in minimal wear. The NOR structure, on the other hand,

concentrates the programming current to the source side corner of the gate oxide, resulting in a decrease in wear-out endurance compared to the NOR structure. The NOR structure is faster than the NAND structure, due to the lower current achieved through tunneling.

The cell architecture of a NOR flash memory, Fig. 2, provides direct access to an individual cell, allowing random access reading and programming. Because of the need for one bit line contact for each two cell group, the dimensions of a NOR array are about 140 percent larger than those of a NAND technology [3]. A typical NAND array stacks 16 floating gate transistors connected in series along with two control transistors. This arrangement eliminates the bit line contacts between the cells as shown in Fig. 3, which allows more transistors to be placed into the die. Operations of both circuits are explained in [4].

3 wordlines and 3 bit lines shown
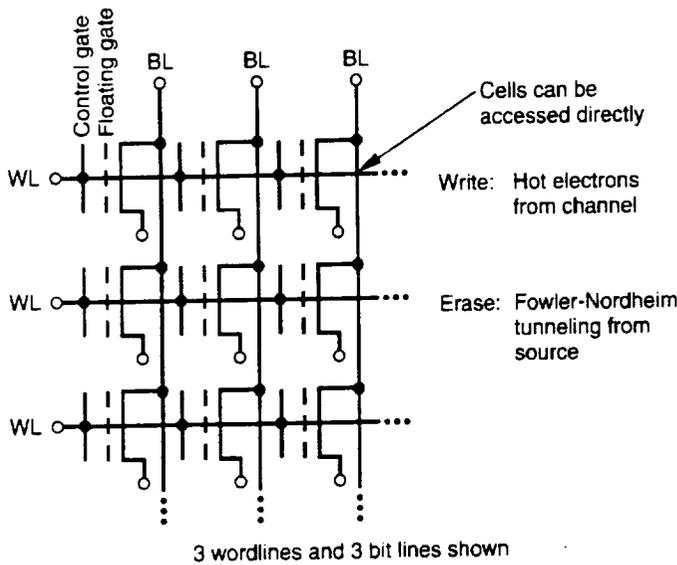
Fig. 2. Cell architecture of a NOR flash memory.

When considering flash memories for space missions, any application that uses intensive read/write cycling will be of the most sensitive to radiation. This mode is important in data recorder or other applications where flash memories are being used in designs that were previously restricted to DRAMs or SRAMs. In such applications, like X2000, there is the concern of wear-out, which occurs after a large number of erase/write cycles. Since the intensive read/write mode will be used in future applications, the robustness of this mode in space environment must be well understood. Most of the radiation liability of flash memories arises from the circuit elements that read and write the floating gates, as the studies below show.

Memory stack height is 16 cells, plus 2 select cells
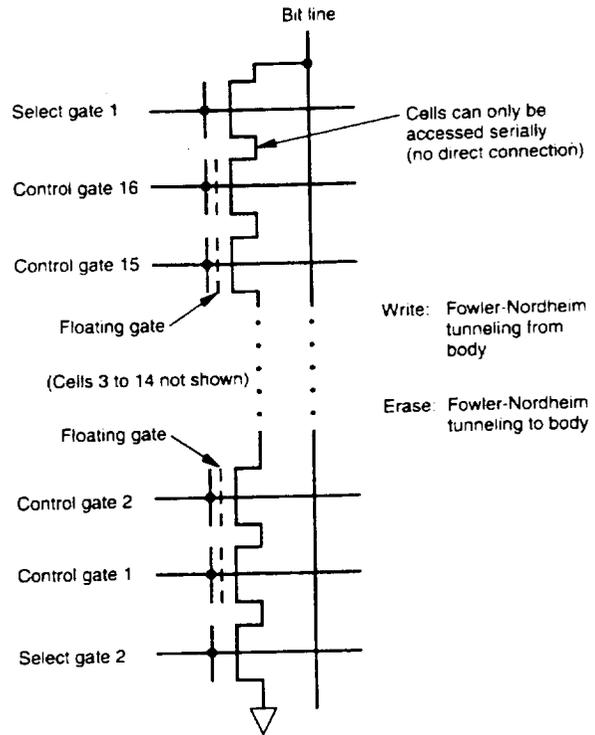
Fig. 3. Cell architecture of a NAND flash memory.

To study the effects of radiation on the floating gate, the UVPROM allows the easiest investigation. This arises from the UVPROM's simple on-board circuitry. The structure of the UVPROM's cell is similar to other NVMs, like the EEPROM or flash memories. UVPROMs are erased only by exposure to radiation [5]. UV radiation is specified by the manufacturer to erase the device. Electron-hole pairs are generated in all areas of the circuit when ionizing radiation interacts with microelectronic circuits [5]. In the FAMOS cell of a UVPROM, some of the holes may interact with the floating gate of the cell to reduce its stored charge [5]. Electrons may also be removed from the gate by direct radiation interaction.

An energetic charged particle generates electron-hole pairs according to its Linear Energy Transfer (LET = stopping power / target density), and the amount of charge removed from the floating gate depends on the LET and the proximity of the trajectory to the gate. One would expect different radiation types to have similar but not identical effects because of different recombination rates and mechanisms [6]. The effect of any exposure to ionizing radiation is the partial removal of charge from the floating gate. This, in turn, reflects the amount of exposure received. By measuring the amount or radiation require to remove all of the charge from the floating gate, the sensitive volume of the oxide which makes up the

40

collection region for erasure surrounding the floating gate of the FAMOS cell can be determined.

2. EXPERIMENTAL PROCEDURES

A. Flash System Analysis

I. Devices Selected for Study

Flash memories from two manufacturers were selected for wear-out. They are listed in Table I. The charge pump effects were studied using the Samsung devices. The wear-out endurance of the Intel parts was studied to support this correlation.

TABLE I
DEVICES SELECTED FOR STUDY

DEVICE	SIZE	MFG	TECHNOLOGY
28F128	128-Mb	Intel	NOR multilevel flash
KM29U128	128-Mb	Samsung	NAND flash
AMD27C64	64-kb	AMD	UVRPOM

Electrical tests after irradiation level were made using an Advantest 3342 test system. The biased circuits follow the DC standby current characteristics of Intel and Samsung specifications. Tests include evaluation of the write operation that requires initial erasure of the flash memory. TID tests were performed at the JPL cobalt-60 room-type irradiator. The devices under test were irradiated under standby power conditions at a dose rate of 25 rad (Si)/s.

II. Charge pump circuit

The typical charge pump circuit is based on the circuit proposed by Dickson in 1976 [7]. The Samsung charge pump circuit uses MOS transistors to accomplish the diode function, as shown in Fig. 4. The circuit operates by charging the coupling capacitors successively each half-clock cycle. The transistors, functioning as forward bias diodes, hold an incrementally larger voltage at each step. The open-circuit voltage can be calculated with the following equation:

$$V_{out} = V_{cc} - |V_{te}| + N (\propto V_{cc} - |V_{te}|) (1)$$

Where V_{te} is the effective V_{th} of the pass transistor, $\propto$ is close to 1 in practical cases, N is the number of stages.

So, even in standby mode, the charge pump is fully engaged. The generated output voltage is directly proportional to the number of charge pump stages and, therefore, it can be generated even at a very low supply voltage (< 2V).
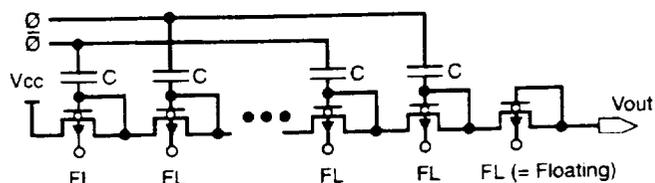
Fig. 4. Samsung flash memory charge pump circuit from [7].

III. Set-Up for Charge Pump Measurements

The reason for suspecting charge pump degradation in wear-out is obvious. Data from previous papers had shown that the degradation of the internal charge pump affected the write and erase capabilities of flash memories [1, 2, 4]. The TID failure level of the Intel 28F016SV flash memories with two power sources (by passing the internal charge pump circuits) was 100 krad(Si), compared to the 25 krad(Si) with the 5V-only power option [4]. Direct measurements of charge pump voltage were accomplished through a probing contact at the die. The probe was connected to a digital multi-meter to read the charge pump voltage of the flash memory arrays. In this study, a Samsung flash memory device was de-lidded to expose the probing pads.

IV. Wear-out Test Procedure

Current test programs were modified to exercise each selected block repeatedly. The test program erases the contents of the flash memory devices (verification of "1" to ensure that the erase process was successful), writes "0" (inversion of data) and then reads "0" (verification of writing data) in all locations of a selected block. The cycle repeats until the test program loop ends or is cancelled.

Two Intel devices were used to study the effects of TID on wear-out of the erase capability. Intel 28F128 devices that operate after being irradiated at 7krad(Si) were then cycled until they failed to erase. Two Samsung devices were tested for wear-out after 9 krad(Si). The selected Samsung device was cycled until it failed to write.

B. Floating gate studies

The devices used in this study were AMD27C64 series UVPROMs consisting of 65,536 FAMOS cells in an 8192x8 bit format. UV radiation can erase the device in part due to a quartz lens encased in the ceramic dip directly over the cells. The normal commercial use of this device is as a read only memory. The UVPROM is exposed to low energy (<8 eV) ultraviolet radiation, which removes electrons from the floating gate, if erasure is desired.

41

The floating gate has been shown to be the most robust subsystem of floating gate memories, but as floating gate memories become scaled, they floating gate may become the foremost radiation liability. The UVPROM were programmed to capacity and interrogated after exposure to radiation. The amount of erasure is measured in terms of equivalent UV radiation. In this manner, the effects of different radiation type, LETs, dose rates etc, can be investigated by only pin interrogation. This is described in detail in [8]. Since the radiation response can be precisely measured and calibrated, floating gate devices can be successfully used as dosimeters. This has been shown on the MPTB satellite and is described in detail in [8].

3. RESULTS

A. Flash Charge Pump Degradation

Fig. 5 shows that the charge pump voltage is constant until the memory device reaches 6 krad(Si) where the voltage drops about 100 mV (from 12.53V to 12.43V). After 7 krad(Si), the measured voltage decreases to 11.8V, but the device still can be erased and written with the new pattern. At the next level of 8 krad(Si), it fails catastrophically during the erasure of 1,024 blocks of memory cells, and the charge pump voltage drops sharply to 6.7V. Direct probing has shown that TID stops the charge pump from providing the expected output voltage for proper erase operation. The dose level at which this device fails is within the distribution of typical radiation failure levels for this device type [1].
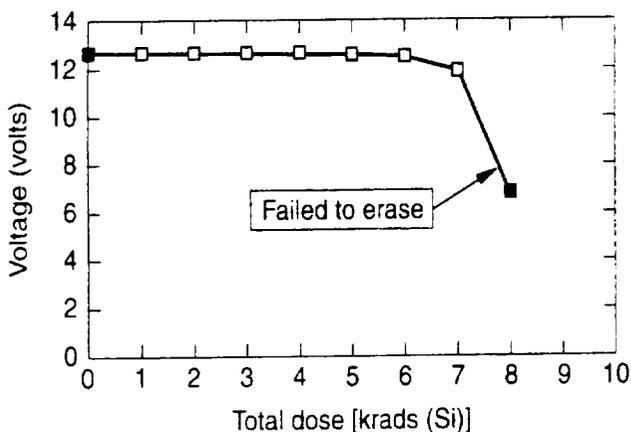
Fig. 5. Charge pump voltage versus total dose.

B. TID Effect on Wear-out

Wear-out is the endurance of flash memory erase/write functions, and it is usually specified by manufacturers as a given number of erase/write cycles. As described earlier, there is some concern about wear-out in solid-state recorder applications that would require many read/write sequences. However, the selected block must

be erased before any data can be written into the arrays of a flash memory cell. If the block cannot be erased or is only partially erased the programmed data will contain errors.

Figs. 6 and 7 show wear-out results. After a dose of 7 krad(Si), both Intel devices failed to erase the selected block at about 20,000 cycles compared to 95,000 cycles for an unirradiated part as shown in Fig. 6. Erase errors are permanent failures to remove electrons from the floating gates and registered as "0"s by the sense amplifier circuitry in the read operation. In this study, both irradiated parts were tested with the maximum allowed erase time as specified by Intel (5 seconds per block erasure). Write errors of "0" data are failures to place correct amount of electrons to the floating gates and registered as "1"s during the read operation. It is also obvious that TID accelerates the erase/write wear-out effects of both Intel and Samsung flash memories. Since erase/write are the only operations that use the 12.5V, the degradation must result from the decay of the charge pump circuitry and previous studies have shown that the read operation is unaffected up to 50 krad(Si) [2]. Fig. 8 shows another effect of the charge pump degradation. Radiation lowers the charge pump voltage thus reduces the tunnel current removing the electrons more slowly.
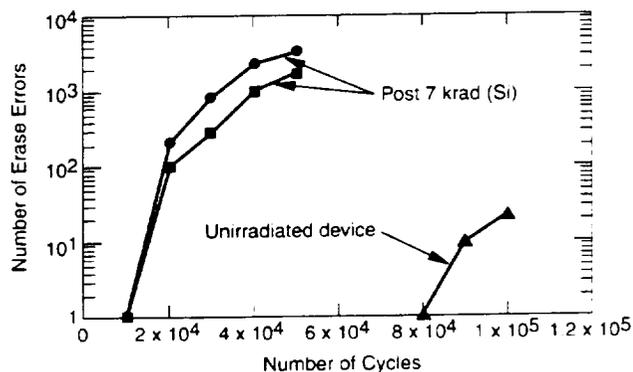
Fig. 6. TID effects on wear-out of Intel devices. Erase errors are failures to remove charge from floating gates.
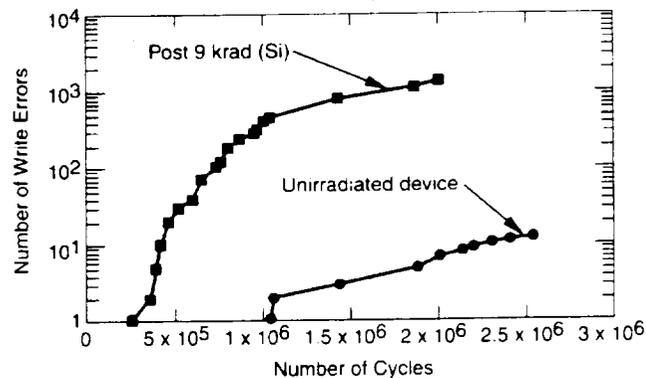
Fig. 7. TID effect on write wear-out of a Samsung device. Write errors of "0" data are failures to place correct amount of electrons to the floating gates.
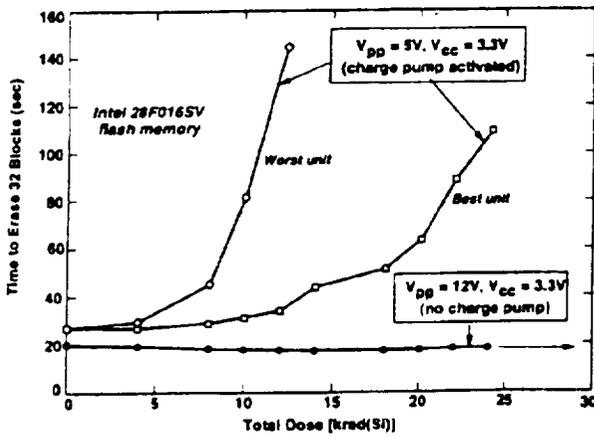
Fig. 8. Effect of dose on timing. The diminished charge pump take longer to induce tunneling in erasure.

C. UVPROM Cell Erasure Results

The plots the fraction of cell erasures as a function of doses for various radiation types are shown in Fig. 9. Fig. 9 shows the result of selected live readout methods. Shown are UV, 50 MeV protons and 1 GeV Argon ions. These erasure curves were directly measured during experiments at accelerators. Fig. 10 shows the fraction of charge on the gate as a function of dose.
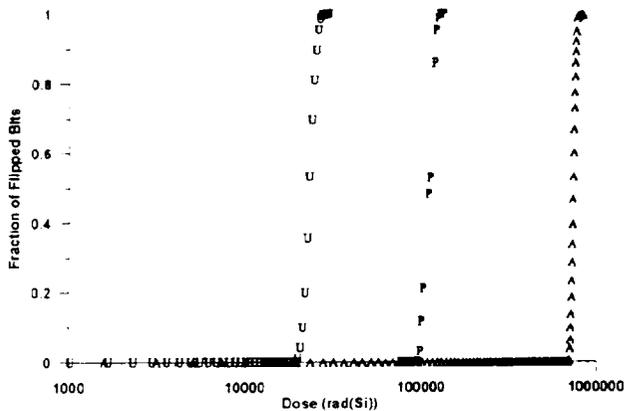
Fig. 9. The live readout erasure response of the device to various radiation types. The U point protector is UV, the P is 50 MeV protons, and A is 1 GeV Argon ions.

From Figs. 9 and 10, the total amount of radiation require to erase a device can be measured. The power law behavior is most likely due to varying sensitivity to dose as a function of dose. The field of the floating gate will decrease with dose, which leads to an exponential relationship, i.e., $E = E_0 e^{\frac{-Dose}{Dose_0}}$, where E_0 is the initial field and $Dose_0$ is a constant. This may explain why higher LET particles are less effective at erasing the

device. So there must be two new erasure mechanisms that are shown here. The rate of erasure is based on dose and also LET, revealing that the oxide around the floating gate must contribute to the erasure mechanism. It is impossible to differentiate between these effects using only the output from the pins of the device.
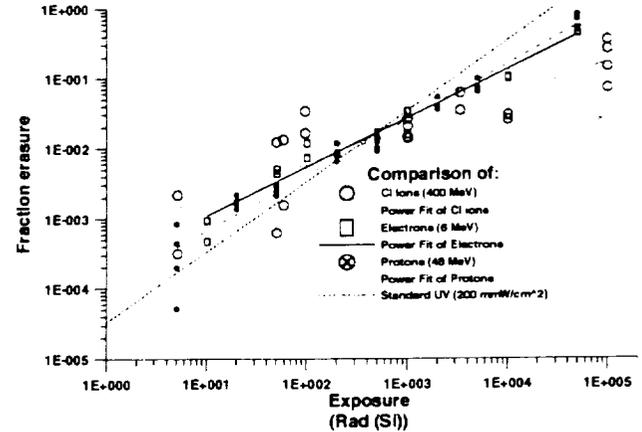
Fig. 10. The fraction of charge left on the floating gate as a function of radiation.

C. UVPROM Sensitive Volume Calculation

By looking at the relationship in Fig. 10, one can start to see how much the oxide affects the erasure process. The relationship is an exponential dependence of erasure efficiency on LET. Not surprisingly, higher LET radiation experiences higher recombination and thus erasure rates are affected. A more obvious statement of this can be seen through the calculation of the effective sensitive volume thickness which is given in detail in [6]:

$$t_{eff} = \frac{1.8 * 10^5 \, eV}{\rho D_{sat} A_{FG}}. \tag{1}$$

The relationship between t_{eff} and LET is shown in Fig. 11. The relationship is again exponential, revealing that the effective thickness of the collecting oxide decreases with increasing LET. The reason for this inverse effect may be due to the charge generation profile of higher LET particles.

D. UVPROM Dosimetry on the MPTB Satellite

One of the important applications of floating gate devices is the measurement of absorbed dose, or dosimetry. Two AMD27C64 UVPROMs were placed in the MPTB experiment as prototype dosimeters. The results are shown in Fig. 12. The first 1200 orbits of the experiment are shown. The other dosimeters, A8, B3, B8, and C4, are shown in solid for comparison. The deviation around orbit 800 is postulated to be saturation in the RADFETs.

43
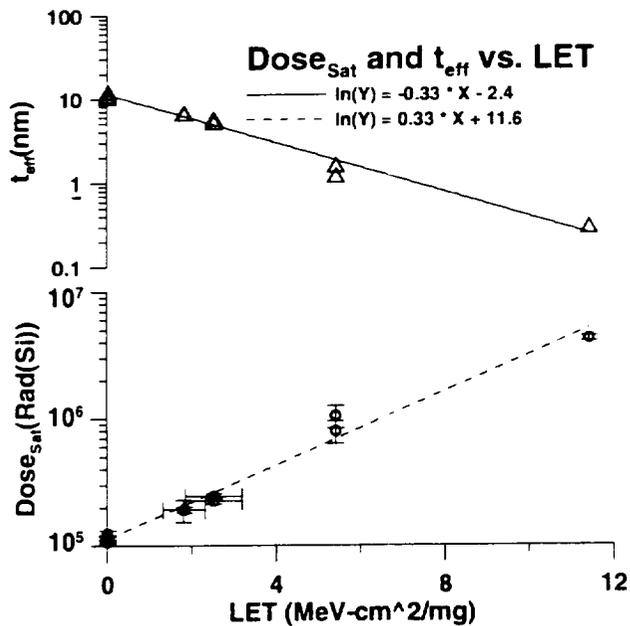
Fig. 11. Total dose required to erase the device versus LET of various radiation.
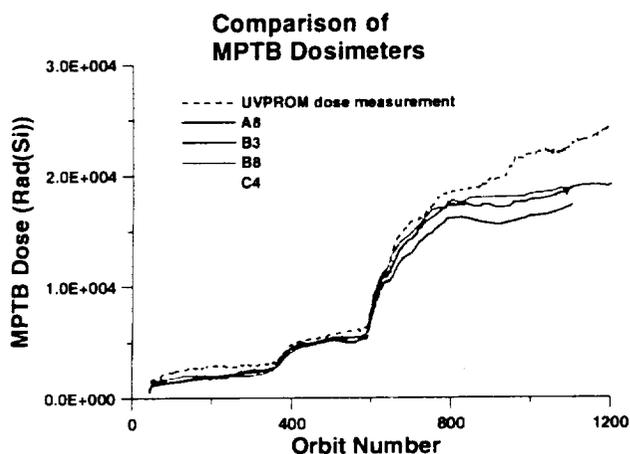
Fig. 12. Dose measured on the MPTB experiment..

IV. DISCUSSION

Several previous studies of radiation effects on EPROM and EEPROM floating gates have reported that data integrity could stand up to 10^5 rad(Si). TID in the range of 10^3 to 10^5 rad(Si) is not expected to upset the data in the memory transistor [1]. SEE studies also concluded that radiation damages the peripheral circuitry before affecting the cell [1, 2]. The total dose failure levels of a 256K EEPROM are limited to values of 10 to 30 krad(Si) due to loss of drive capability of the peripheral circuitry, not charge loss from the memory transistors [1]. The most sensitive control circuitry in flash memory devices is the internal charge pump circuit. Any leakage current within the internal charge pump circuit will lower the output voltage. Serious leakage currents can be conducted

by the capacitors coupled in parallel with two non-interleaving clock signals. The leakage current increases with TID. Any stage-to-stage increases in leakage would also reduce the charge pump output voltage [3]. For NAND devices it should be noted that the highest voltage is needed for erasing (20V), then programming (18V), then reading , this is the order in which failures are seen in the TID tests on the Samsung 128Mb flash memory [2].

Charge injection through the dielectric by tunneling causes hole trapping at the interface. The trapped positive charge induced parasitic leakage current. The floating-well charge pump circuits can also generate substrate current. The resulting currents can clamp the charge pump output voltage to low values, making the flash memory inoperable.

V. CONCLUSIONS

Of the conclusions that can be drawn from this work, the most outstanding is that TID directly affects the output of the internal charge pump. Radiation also reduces the endurance of flash memories due to operations that depend on charge pump output voltages. Thus there is a correlation of charge pump degradation and the number of erase/write cycles. This relationship needs more quantitative investigation. Future studies will quantify the effects of TID on wear-out.

REFERENCES

[1] D.N. Nguyen, C.I. Lee, and A.H. Johnston, "Total ionizing dose effects on flash memories," *1998 IEEE Radiation Effects Data Workshop Record*, p.100.

[2] D.N. Nguyen, S.M. Guertin, G.M. Swift, and A.H. Johnston, "Radiation effects on advanced flash memories." *IEEE Trans. Nucl. Sci.*, vol. 46, no. 6, pp. 1744-1750, Dec. 1999.

[3] W. Brown, *Nonvolatile Semiconductor Memory Technology*, IEEE Press, 1998.

[4] H.R. Schwartz, D.K. Nichols and A.J. Johnston, "Single event upset in flash memories." *IEEE Trans. Nucl. Sci.*, vol. 44, pp.2315-2323, Dec. 1997.

[5] B. Prince and G. Due-Gundersen, *Semiconductor Memories*, New York: John Wiley and Sons, 1983.

[6] P. McNulty, L. Scheick, and D. Roth, " Measurement of the Effective Sensitive Volume of FAMOS Cells of an Ultraviolet Erasable Programmable Read Only Memory." *IEEE Trans. Nucl. Sci.*, To be published.

[7] J. Dickson, "On-chip high-voltage generation in MNOS integrated circuits using an improved voltage multiplier technique." *IEEE J. Solid State Circuits*, vol. SC-11, pp. 374-378, June 1976.

[8] L. Scheick, P. McNulty, and D. Roth, "Dosimetry based on the erasure of the floating gates in the natural radiation environments in space." *IEEE Trans. Nucl. Sci.*, vol. 45, no. 6, pp. 2681-2688, 1998.

Raytheon

Total Dose Radiation Response and High Temperature Imprint Characteristics of Chalcogenide Memory Elements

Steve Bernacki, Raytheon Sudbury MA
Ken Hunt, Scott Tyson, AFRL Albuquerque NM
Steve Hudgens, Boil Pashmakov, Wally Czubatyj, Ovonyx Troy MI

July 27, 2000

NSREC Reno NV

NSREC 2000 Rev 3 page 1 printed 6/22/00

Raytheon

C-RAM Memory Element Radiation Testing

Agenda

- Motivation

- Chalcogenide Technology Primer

- Temperature and Total Dose Testing Methodology

- Test Results and Discussion

- Conclusions

NSREC 2000 Rev 3 page 2 printed 6/22/00

C-RAM Memory Element
Radiation Testing

Raytheon

Motivation

* Commercial world needs nonvolatile memory
  - Flash is current baseline, but has limited speed and endurance

* Space programs have additional requirements
  - Nondestructive Read Out (NDRO)
  - Radiation hardened to both total dose and single event environments

* Three technologies have been investigated with limited success
  - Silicon-Oxide-Nitride-Oxide-Silicon (SONOS) EEPROM
  - Ferroelectric (FRAM, FeRAM)
  - Magnetic (MRAM, GMRAM, MAGRAM)

* Recent emerging technology appears promising
  - Chalcogenide (C-RAM)

NSREC 2000 Rev 3 page 3 printed 6/22/00

C-RAM Memory Element
Radiation Testing

Raytheon

Chalcogenide Memory Technology Primer

* Basic memory element consists of a variable resistance thin film deposited before metalization in semiconductor process
  - $Ge_2Sb_2Te_5$ (Group VI elements (Te) are the chalcogens)
  - Can be programmed between polycrystalline or amorphous phases
  - Resistance changes by at least two orders of magnitude
    * High resistance in amorphous phase
    * Low resistance in polycrystalline phase

* Reversible phase change affected by heating and rapid quenching
  - In this case, by resistive heating with nanosecond pulses

* Thin film sputtered at room temperature from compositional targets
  - Fully compatible with CMOS processing for monolithic integration

NSREC 2000 Rev 3 page 4 printed 6/22/00

46

C-RAM Memory Element
Raytheon

Radiation Testing

Cross Section of Test Structure

- First attempt to integrate chalcogenide with silicon
  - Micron Semiconductor devices *circa* 1996
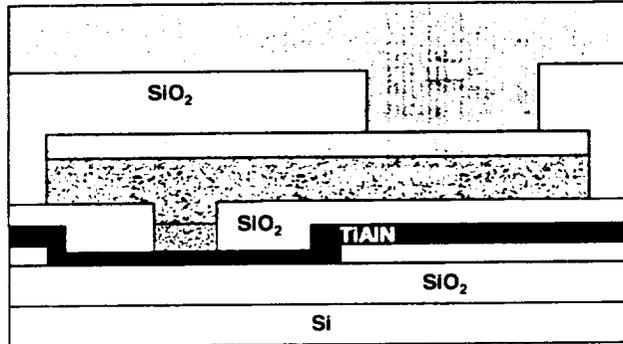
NSREC 2000 Rev 3 page 5 printed 6/22/00

C-RAM Memory Element
Raytheon

Radiation Testing

Programming Characteristics
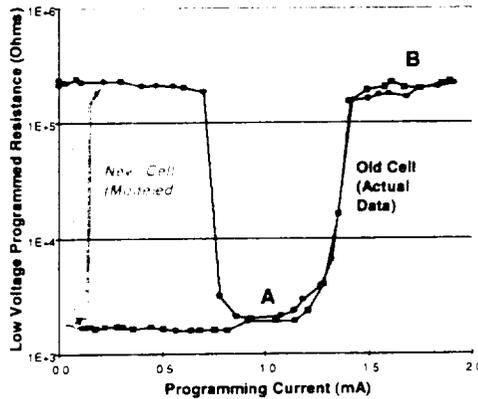
– Application of 1.0 mA (**A**) will nucleate conductive crystallites

– Application of 1.5 mA (**B**) will melt small volume of material, rapid quenching locks in disordered low conductive state

NSREC 2000 Rev 3 page 6 printed 6/22/00

C-RAM Memory Element
Radiation Testing

Chalcogenide Technology Commercial Base

- Technology patented by Energy Conversion Devices
  - Licensed by Ovonyx, joint venture of ECD and Tyler Lowrey

- Identical chalcogenide material as used for CD-RW and DVD-RAM
  - Both optical reflectivity and resistivity affected by free carriers

- Lockheed Martin, Ovonyx developing rad hard memory
  - Present work funded under AFRL contract with MRC

- Intel and Ovonyx developing commercial high density memory

- www.ovonyx.com for everything you ever wanted to know

NSREC 2000 Rev 3 page 7 printed 6/22/00

C-RAM Memory Element
Radiation Testing

Test Circuit Schematic
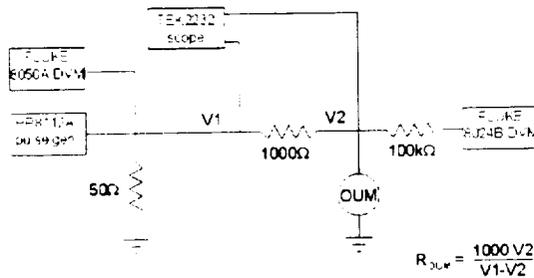
$$R_{DUT} = \frac{1000 \, V2}{V1-V2}$$

NSREC 2000 Rev 3 page 8 printed 6/22/00

C-RAM Memory Element Radiation Testing

Raytheon

Test Circuit Waveforms

Switch from LO to HI   Switch from HI to LO

V2 voltage, mV
600 –
500 –
400 –
300 –
200 –
100 –
0 –

Time, 200ns per division
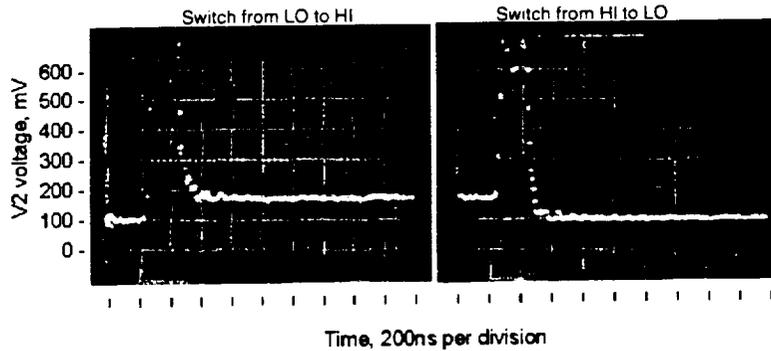
LO state   V1 = 176mV; V2 = 102mV; Roum = 1.48kΩ
HI state: V1 = 177mV; V2 = 176mV; Roum = 176kΩ

- **Confirms Reversible Switching Between States**

C-RAM Memory Element Radiation Testing

Raytheon

Testing Methodology

- Characterize baseline functionality of all pins
  - Functionality: ability to write and read data before and after experiment
  - Retention: ability to retain data throughout experiment

- Program memory elements into alternating LO and HI states

- Imprint data at 125°C, 2 hours, air, electrodes shorted

- Measure retention, functionality, program elements again

- Expose sample with Co-60 irradiation

- Measure retention, functionality, program elements again

C-RAM Memory Element
Raytheon

Initial Baseline Functionality
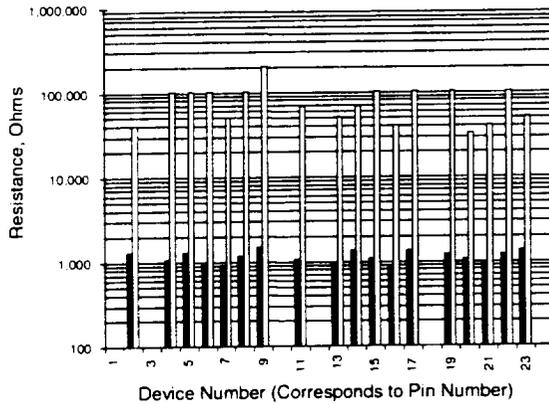
- Odd numbered pins programmed HI and even numbered pins programmed LO for subsequent experiments

NSREC 2000 Rev 3 page 11 printed 6/22/00

C-RAM Memory Element
Radiation Testing
Raytheon

Functionality All Pins vs Temperature and Total Dose
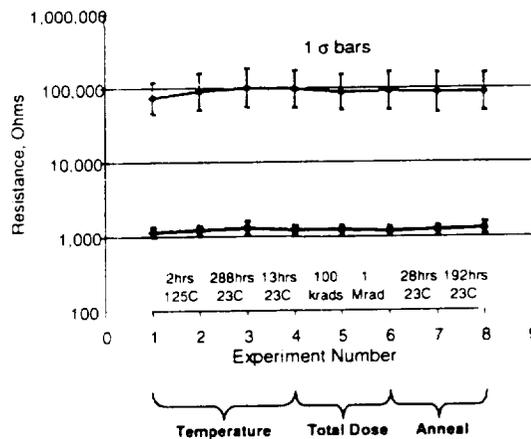
- No observed functionality effects due to temperature or total dose

NSREC 2000 Rev 3 page 12 printed 6/22/00

C-RAM Memory Element Radiation Testing

Raytheon

Functionality Pins 13 and 14 vs Temperature and Total Dose
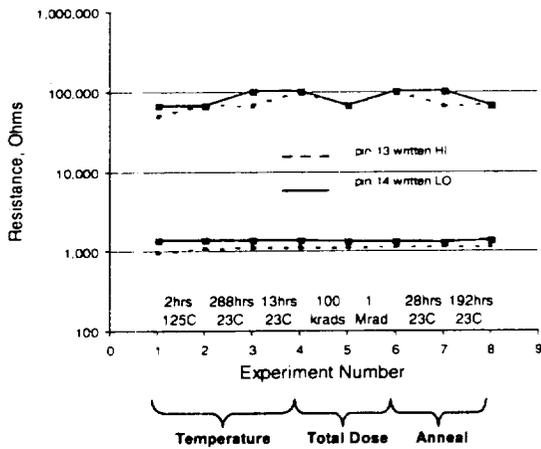
- No observed functionality effects due to programmed state (imprint)

NSREC 2000 Rev 3 page 13 printed 6/22/00

C-RAM Memory Element Radiation Testing

Raytheon

Retention Pins 13 and 14 vs Temperature and Total Dose
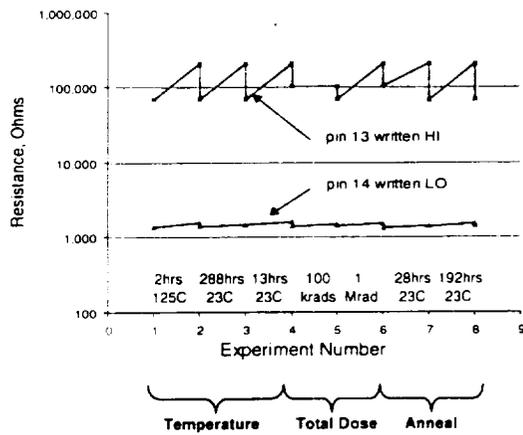
- No observed retention effects due to temperature or total dose

NSREC 2000 Rev 3 page 14 printed 6/22/00

C-RAM Memory Element
Radiation Testing

Raytheon

Conclusions

- Programmable memory elements work as advertised
  - Reversible, fast switching up to two orders of magnitude resistance

- Memory element properties unaffected by environments
  - Functionality and retention
  - 125°C temperature for 2 hours
  - Total dose irradiation to 1Mrad(Si)
  - Elements programmed in either HI or LO state

- Future plans
  - Neutron testing in progress
  - Total dose, neutron, and SEU testing on new structures from Lockheed Martin

NSREC 2000 Rev 3 page 15 printed 6/22/00

C-RAM Memory Element
Radiation Testing

Raytheon

Acknowledgements

- This work was funded by AFRL Albuquerque through Mission Research Corporation

- Bob Pugh, AFRL, for program encouragement and support

- Tyler Lowrey and Ed Spall, Ovonyx, and Stan Ovshinsky, Energy Conversion Devices, for technical collaboration

NSREC 2000 Rev 3 page 16 printed 6/22/00

52

SONOS Nonvolatile Semiconductor Memories for Space and Military Applications

Dennis A. Adams[1], David Mavis[2], James R. Murray[3], and Marvin H. White[4]

[1] Northrop Grumman Corporation
Baltimore, MD
dennis_a_adams@md.northgrum.com

[3] Sandia National Laboratories
Albuquerque, NM
jrmurra@sandia.gov

[2] Mission Research Corporation
Albuquerque, NM
dgmavis@rt66.com

[4] Lehigh University
Bethlehem, PA
mhw0@lehigh.edu

Abstract—Silicon-Oxide-Nitride-Oxide-Silicon (SONOS) based nonvolatile memory has emerged as the most mature nonvolatile semiconductor memory (NVSM) currently in use for space applications. SONOS 64k EEPROMs have been flying in numerous satellite applications since 1992 with a 256k EEPROM version of this part qualified for space applications in 2000. This paper will summarize the production and development status of a family of SONOS-based devices (EEPROMs, FPGAs, Controller ASICs) currently being manufactured at the Northrop Grumman Corporation (NGC) Advanced Technology Laboratories in Baltimore, Maryland.

TABLE OF CONTENTS

1. INTRODUCTION

Northrop Grumman Corporation (NGC) (formerly Westinghouse Electric Corporation) has been involved in nitride-based nonvolatile memory for space since the middle 1970's [1-3]. The Silicon-Oxide-Nitride-Oxide-Silicon (SONOS) process is a radiation hard (300 krad total dose) 1.2 μm CMOS technology for EEPROM applications. This technology has been the result of a joint collaboration with Sandia National Laboratories and Lehigh University.

SONOS transistors are built on a stack of 15 A thermal oxide, 150 A silicon nitride and 40 A blocking oxide

Figure 1 - SONOS Basic Mechanisms
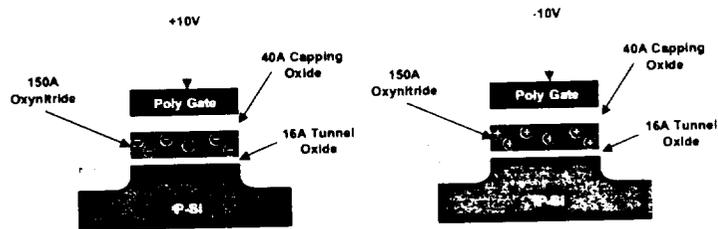
- +10 volt programming results in trapped electrons in SONOS dielectric stack
- -10 volt programming results in trapped holes in SONOS dielectric stack
- Data loss caused by emission of trapped charge from SONOS stack

under a phosphorous doped polysilicon gate. This stack is then programmed with either +10 V or 10 V. These two

bias conditions result net trapped negative or positive charge, respectively as shown in Figure 1.

The SONOS process offers dramatic radiation hardness improvements over conventional floating gate EEPROM technology. The key reason for this is in the difference in how charge is stored for these two technologies. SONOS technology relies on trapped charge that is stored in a nitride dielectric that is not easily removed in a total ionizing dose environment. Floating gate memories store charge in conducting floating gates that are separated from the silicon substrate by thin (<100 A) tunnel oxides. Total ionizing dose radiation damages these tunnel oxides leading to a more rapid charge loss for floating gate devices. Heavy ion strikes from galactic or solar radiation can similarly discharge floating gate devices. Both of these effects severely limit the radiation hardness of floating gate devices and are not observed with SONOS devices.

NGC 64k / 256k EEPROMs are specified to withstand 300 krads (Si) total ionizing dose [testing to 450 krads(Si)

with 1 week biased anneal according to MIL-STD-883 Test Method 1019.5] SONOS 256k EEPROMs have been observed to retain data after total dose exposure of over 800 krads(Si). This level of hardness is not possible with today's floating gate EEPROMs.

Trapped charge in SONOS transistors is stored in traps within the nitride dielectric. Charge loss for SONOS devices is dominated by well behaved thermal emission effects that are linear with the logarithm of time. Typical SONOS memory retention is in the 10 to 100 year range depending on the pulse width of the programming voltage. (An example of typical SONOS transistor memory retention is shown in Figure 2.)

SONOS Retention Characteristic
Vp-+.10v, Precond. +-10v 10s

Retention Time (Sec.)
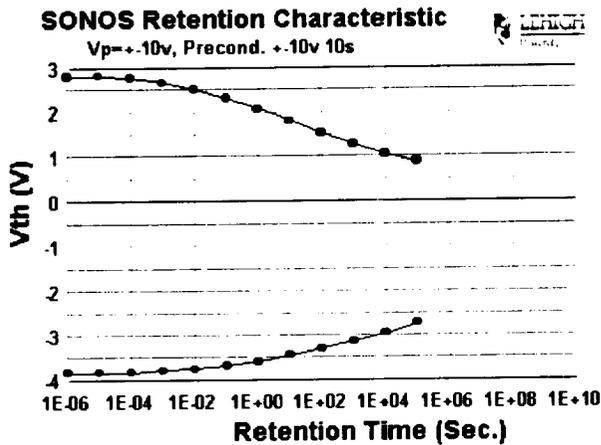
Figure 2 – Typical SONOS transistor threshold voltage retention decay data showing >10 year retention (1E8.5 sec).

Historically, SONOS nonvolatile memories have been conservatively specified for a maximum of 10,000 (10V) erase-write cycles for 10-year memory retention. Transistor level memory retention data indicates negligible change in memory retention at this level of endurance cycling. At 100,000 cycles transistor level tests indicate parts would pass 10-year memory retention requirements. However, a full study to evaluate increasing endurance cycling from 10,000 to 100,000 cycles has not been done. Recent work by Lehigh University indicates future SONOS nonvolatile memories with 10^6 to 10^7 erase-write cycles are possible.

Lehigh University has also performed extensive characterization and process development of scaling SONOS for low voltage applications [4,5]. Their work has resulted in SONOS stacks which can be programmed with voltages as low as 5 V (Figure 3). Lehigh work in the area of deuterium annealing has demonstrated SONOS endurance improvements by over an order of magnitude. NGC is currently doing work to incorporate these scaled SONOS and deuterium anneal results into future submicron nonvolatile memory applications.
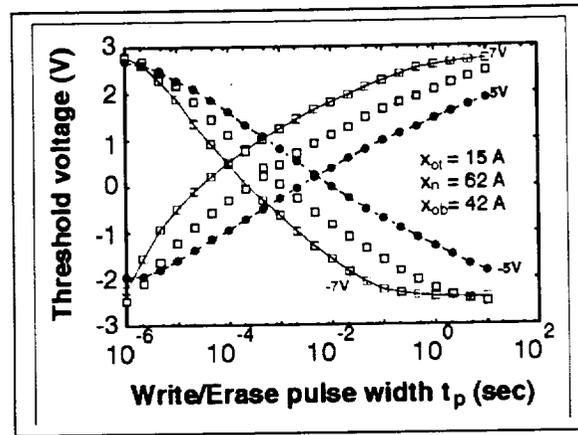
Figure 3 – Lehigh data showing 5 V programmable SONOS stacks.

2. 64k /256k /1 Mb EEPROMs

The NGC 64k (8K x 8 bits) EEPROM is in use in over 20 satellite applications and has demonstrated radiation hardness in excess of 300 krads. Over 6000 space qualified 64k EEPROMs have been shipped to date. The 256k (32K x 8 bits) EEPROM device uses the same memory cell and peripheral circuitry as the 64k EEPROM with qualification in March 2000. The 256k EEPROM has been demonstrated to have negligible increase in standby current up to 1.5 Mrads with no loss in functionality (Figure 4). A table of the key specifications for these parts is included in Table 1.
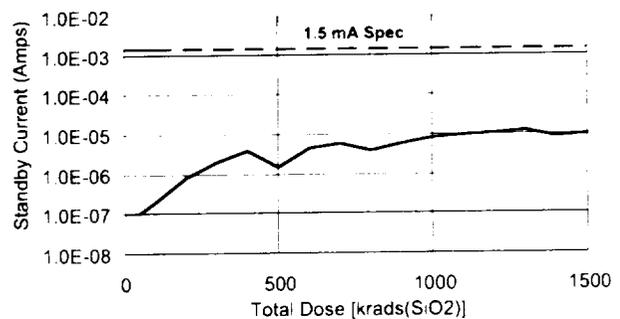
Figure 4 - 256k CMOS/SONOS EEPROM standby current <15 uA at 1.5 Mrad(Si) [Specification limit = 1 mA.]

Table 1 – 64k (256k) EEPROM Specifications

Total ionizing dose:	300 krads(Si)
Latch-up:	NONE
Transient upset: - Logic - Memory	>5E7 rads(Si)/sec >1E12 rads(Si)/sec
Single Event Upset: - SEU during READ - SEU in address/data latches - Permanent damage	LET_{th}=60 MeV/mg/cm**2 LET_{th}=35 MeV/mg/cm**2 Atomic # ≥ Krypton
Memory retention: - STD (10 msec) - ROM (100 msec)	10 years 100 years
Endurance:	10,000 cycles
Access time:	250 nsec
Temperature range:	-55 C to +125 C

The 1M (128K x 8 bits) Rad-Hard EEPROM is being designed at Sandia National Laboratories for fabrication in the NGC 0.8 um double level metal CMOS/SONOS process. Each memory cell contains one SONOS transistor. This transistor is compared with a reference SONOS transistor, which is always erased. The difference in their threshold voltages represents the binary value of the bit. For the data in the cell to be read correctly, one or both SONOS transistors must be depletion mode (i.e. threshold voltage less than zero). The specification goals for the design are listed below:

Table 2 – 1 Mb EEPROM specification (projected)

Power Supply	5v or 3.3v (I/O)
Programming Voltage	10v (-5v - +5v) or (-6.7v - +3.3v)
Clear/Write Time	<10ms
Endurance	1e4 Cycles
Read Access Time	<250ns
Retention	10 Yrs (80C) 25 Yrs (ROM Applications)
Temperature	-55C - +125C
Total Dose Radiation	300K Rad(Si)
Transient Radiation	
Logic Upset	>5e7 Rad(Si)/S
Data Upset	>1e12 Rad(Si)/S.
SEU Immunity	>35 MeV-cm2/g
SEU Permanent Damage	>Ar

This part is pin compatible with commercial floating-gate 1M EEPROMs (Hitachi, Atmel, etc.). The block diagram is shown below. Commercial parts are packaged in 32 pin packages. The Rad-Hard 1M EEPROM requires additional pins, so it will be packaged in a higher pin count package (e.g. 36 FP). The additional signals will use the top and bottom pin locations (i.e. pin 1 - VW, pin 18 - CLK, pin 19 - RSTB, pin 36 - PE). This allows a 32 pin commercial EEPROM to be used in a socket designed for the Rad-Hard 1M EEPROM by leaving pins 1, 18, 19, & 36 of the socket open. A block diagram of the 1M EEPROM is shown below:

EEPROM Block Diagram
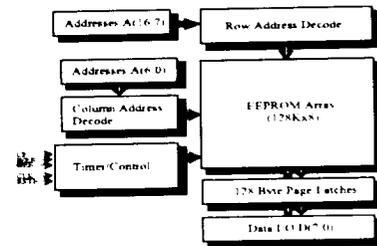
Figure 5 – Block diagram for SNL / NGC 1M EEPROM.

3. CMOS ASIC CONTROLLERs

SONOS technology has been base-lined into radar applications at Northrop Grumman for over 10 years. SONOS EEPROMs are employed in CMOS ASIC transmit/receive module controller ASICs, which have memory retention requirements in excess of 40 years. These devices are a critical part of NGC electronically steered active aperture radar technology. SONOS EEPROM memories on these parts store information on module address, analog delay values and GaAs bias settings. Over 40,000 of these devices have already been delivered. SONOS-based module controller ASIC work is continuing for both current and future NGC radar systems. The next generation controller ASIC is currently being manufactured with 0.8 μm CMOS/SONOS technology.

4. 4K FPGA

A 4K field programmable gate array (FPGA) designed by Mission Research Corporation was demonstrated at the 2000 NSREC and MAPLD conferences [6]. This part uses the NGC CMOS/SONOS triple level metal technology. The FPGA uses SONOS transistor memory elements for programming and is the first nonvolatile, reconfigurable FPGA that is radiation hardened for space applications (Figure 6).

A new tierable and nestable directional routing architecture enables the use of standard radiation-hardened circuit design practices and specifically avoids

pass gates for routing switches. The design has 4000 equivalent gates, operates at clock speeds up to 20 MHz and is hardened to >200 krads(Si) total dose. The FPGA uses SEU immune EEPROM storage circuits as well as SEU immune logic latches. The clock and reset control lines are hardened to an LET > 100 MeV-cm^{**2}/mg.
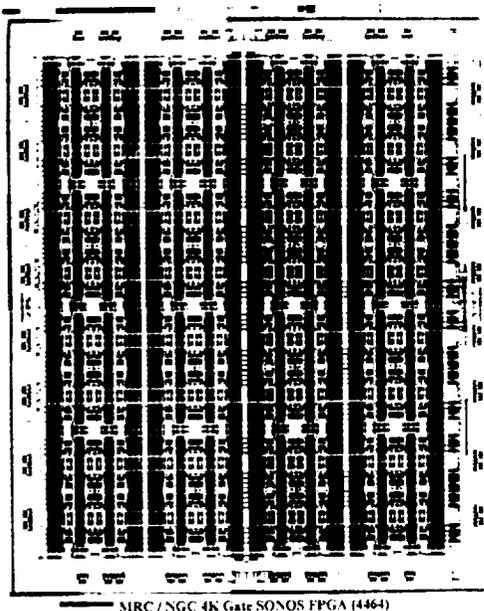

Figure 6 – MRC / NGC rad hard 4K FPGA features reconfigurability using SONOS nonvolatile memory elements.

5. MIXED SIGNAL TILE ARRAYS

SONOS technology is also used in the NGC mixed signal tile array. This device uses 1.2 um BiCMOS technology for analog applications. This product is planned for use in an undersea telecommunication application with a 30-year memory retention requirement and is currently being qualified for production.

The tile array device is similar in concept to a CMOS gate array in that the device can be created by customizing four interconnect photo masks. The actual size of the tile array part can also be modified to use between 1 and 90 "tiles" for further design flexibility. Each tile has an assortment of CMOS/ SONOS bipolar transistors, resistors, capacitors, and Schottky diodes suitable for a wide variation of mixed signal applications (Figure 7).

6. CONCLUSIONS

SONOS nonvolatile memory has a long history as a robust technology that is well suited to meet stringent reliability and radiation hardness requirements for space and avionics applications. A wide range of products are currently in production using SONOS-based nonvolatile memories. NGC, in collaboration with Lehigh University, MRC and Sandia Labs, is continuing research

towards the development of next generation submicron SONOS products.
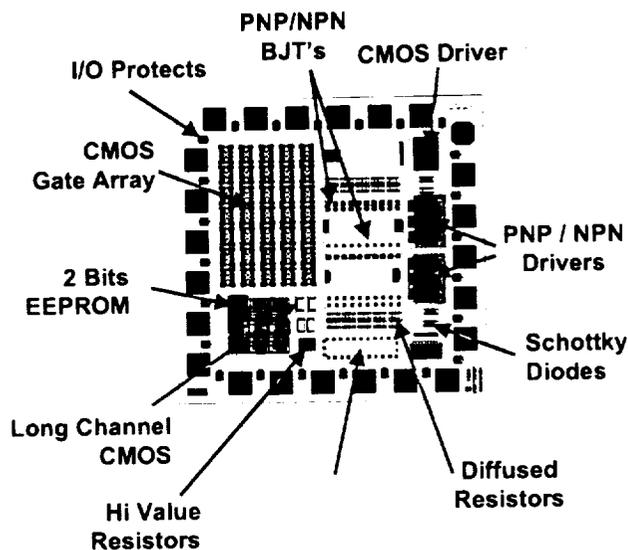
Figure 7 – NGC mixed signal tile array with SONOS EEPROM.

7. ACKNOWLEDGEMENTS

8. REFERENCES

[1] D.Adams, et al, "Radiation Hardened 64k/256k EEPROM Technology, 1996 GOMAC Conference, pp. 188-190.

[2] M.Jacunski, et al, "Radiation Hardened Nonvolatile Memory for Space and Stragtegic Applications", 1992 GOMAC Conference, pp.449-452.

[3] F.Sexton, et al, "SEU and SEL Response of the Westinghouse 64k EEPROM", 1994 IEEE NSREC Data Workshop Proceedings, pp.55-63.

[4] Y. Yang and M.H. White, "Characterization of Low-Voltage Scaled SONOS Memory Transistors", Solid-State Electronics, to be published in 2000.

[5] M. L. French and M. H. White, "Scaling of Multidielectric Nonvolatile SONOS Memory Structures", Solid-State Electronics, 37, 1913 (1994).

[6] D.Mavis, et al, "A Reconfigurable, Nonvolatile, Radiation Hardened Field Programmable Gate Array for Space Applications", 1998 NASA Goddard Radiation Hardened Technology Symposium.

SONOS Nonvolatile Shadow RAMs for Space Applications

Gary F. Derbenwick*, David A. Kamp*, Alan F. Isaacson*, J. Roger Gill** and Stephen Linn**

*Celis Semiconductor Corporation
5475 Mark Dabling Blvd., Suite 102
Colorado Springs, CO 80918
(719) 260-9133; (719) 593-8540 (fax); celis@celis-semi.com

**Simtek Corporation
1465 Kelly Johnson Blvd.
Colorado Springs, CO 80920
(719) 531-9444; (719) 531-9481 (fax); gill@simtek.com

Abstract—Several versions of shadow random access memories have been commercialized. These memories consist of an SRAM shadowed by a nonvolatile memory on a single chip of silicon. Both floating gate and silicon nitride devices have been used for the nonvolatile shadow memory storage elements. Because radiation hardened SRAMs have been proven for space applications and silicon nitride devices have a high tolerance to radiation exposure, shadow RAMs based on the silicon nitride technology can be combined with radiation hardened CMOS to provide rugged nonvolatile semiconductor memories for space applications. These memories have high endurance, fast write times and low power performance because these parameters are determined by the SRAM portion of the memory. The nonvolatile shadow memory is automatically written only on power down. Therefore the endurance limitation of silicon nitride to approximately 10^6 cycles applies only to the number of power down cycles. Nonvolatile retention times of greater than ten years can be obtained at 140°C.

TABLE OF CONTENTS

1. INTRODUCTION

Shadow random access memories (shadow RAMs) consist of a volatile SRAM shadowed bit by bit with a nonvolatile memory. During power up, the nonvolatile shadow memory causes the latch nodes of the SRAM memory cell to charge at different rates, thus setting the SRAM memory in a known state. During power down, the state of the SRAM is stored in the shadow RAM so that the SRAM is restored to the proper state during the next power up. The nonvolatile shadow memory is integrated within each volatile memory cell on a single chip of silicon.

Several versions of shadow RAMs have been commercialized. These consist of SRAMs containing a floating gate nonvolatile transistor, SRAMs containing a pair of silicon nitride capacitors, and SRAMs containing a pair of silicon nitride transistors. Silicon nitride versions are most suited for space applications because floating gate devices are relatively soft to radiation exposure. Further, shadow RAMs using a pair of silicon nitride transistors are best suited since commercial versions of these memories are being produced today, the memory cell size is smaller than when silicon nitride capacitors are used for the storage elements, and during power up the SRAM memory is restored to the true state rather than the complement state.

Radiation hardened SRAMs have previously been used for space applications. Radiation hardened CMOS circuitry is required and radiation hardened design techniques must be used. Previous studies have shown silicon nitride devices to retain data through ionizing radiation doses well in excess of a Megarad.[1,2]

When silicon nitride shadow memories are combined with radiation hardened SRAMs in a shadow RAM architecture, nvRAMs can survive radiation doses typically encountered in earth orbit and in near and deep space missions, including ionizing radiation, proton irradiation and high energy ion bombardment.

2. ARCHITECTURE

Fig. 1 shows a functional block diagram for an nvSRAM. The nvSRAMs are shadow RAMs using two silicon nitride nonvolatile transistors in each memory cell for the nonvolatile storage elements. Under normal operation, the nvSRAM operates the same as an SRAM with similar performance characteristics. During power down, the data in the SRAM is block written into the SONOS shadow RAM in under 10 ms. This is called the store operation. There are three ways to implement the store operation. Fig. 1 shows a software store where circuitry detects a nonsensical address sequence that is interpreted as a command from the system controller that the system is about to be powered down. Alternatively, a hard wired pin on the nvSRAM can be used to supply the power down signal. For automatic operation, enough charge can be stored on a small capacitor on the printed circuit board to enable V_{dd} to be high enough for 10 ms to complete the store operation. In this case, the nvSRAM has automatic power-on-reset circuitry that automatically detects the drop in V_{dd} during power down and commences the store operation.
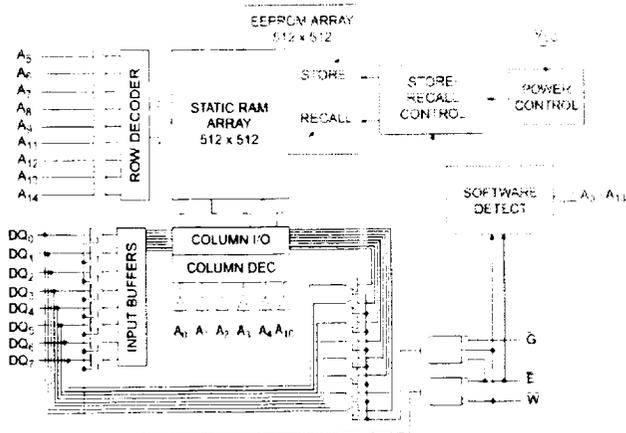
Figure 1. Functional block diagram of an nvSRAM.

Fig. 2 shows a schematic of the memory cell portion of an nvSRAM for a four transistor, two resistor SRAM architecture. The gates of the two SONOS transistors are connected to V_{se} and the drains are connected to the latch nodes of the SRAM cell, dt and dc. One advantage of the nvSRAM architecture is that the higher voltage (8 –15V) required for programming the SONOS transistors is only applied to V_{se}, a polysilicon line, in the memory array.

A timing diagram corresponding to Fig. 2 is shown in Fig. 3. So long as the chip is powered up, the operation stays in the SRAM cycle portion. During power down, the STORE cycle is implemented, and during power up the RECALL cycle is implemented.

While Fig. 2 shows the SRAM cell with two high value resistors, similar cells can be designed using six transistor SRAM architectures. Such memory cells can also be used for radiation hardened version of the nvSRAM memory cell.
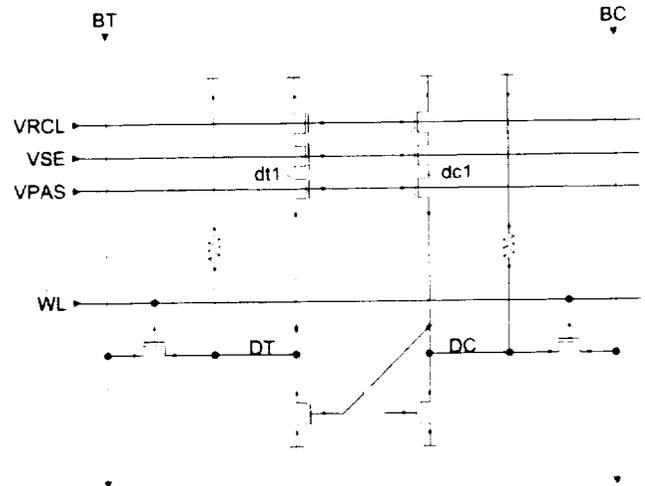
Figure 2. Schematic of an nvSRAM memory cell.
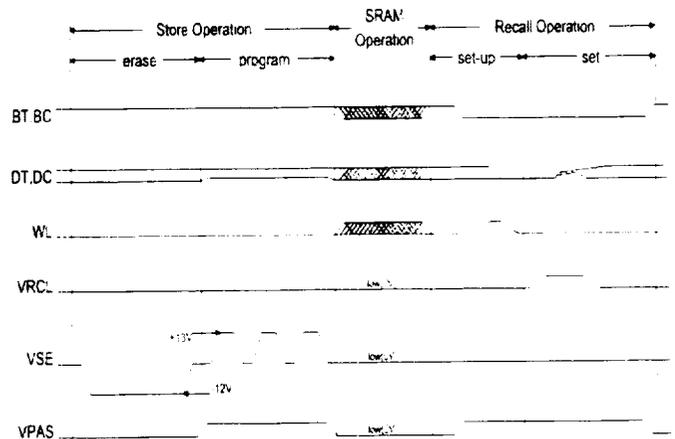
Figure 3. Store/Recall timing diagram for nvSRAM.

3. PERFORMANCE

Table 1 gives other major performance parameters for nvSRAMs.

The SONOS technology has excellent retention performance. The data in Fig. 4 show greater than 10 years retention at 200 °C. Every bit in an nvSRAM can be

electrically screened for retention time to assure retention specifications are met.

Table 1. Major nvSRAM Performance Parameters

Operating Voltage	3V or 5V
Read Cycle Time	25ns
Write Cycle Time	25ns
Endurance	∞
Retention (125 °C)	> 10 yrs
Store Time	10ms
Power Down Endurance	10^6
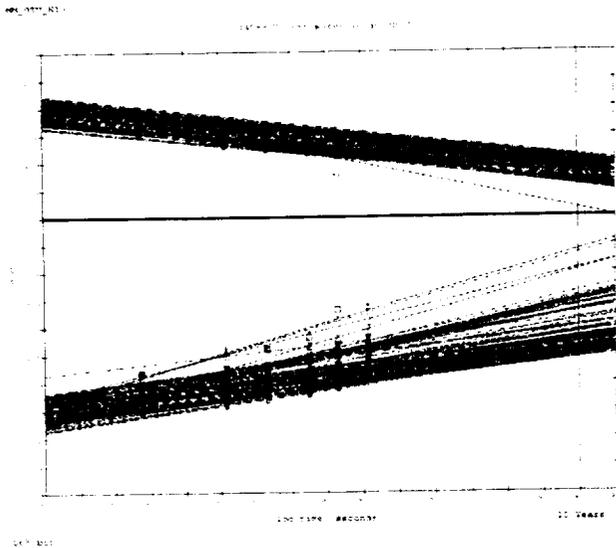
Figure 4. Decay with log time of erased and programmed states of SONOS storage devices.

4. RELIABILITY

The failure rate of commercial nvSRAMs versus year of production are shown in Fig. 5. Total failure rates under 20 FITs are measured, consistent with leading edge volatile semiconductor memories and typically better than reprogrammable nonvolatile semiconductor memories.

5. RADIATION HARDNESS

Radiation resistance of SONOS devices and SONOS memories has been extensively measured. Memory window collapse is typically observed at approximately 10 Mrad ionizing radiation dose.[1,2] This exceeds the ionizing radiation hardness of most CMOS circuitry.

SONOS devices have been shown to be immune to single event upset (SEU) due to high energy ion exposure, with the exception that there is a small, but non-zero, probability that a silicon nitride gate can short if an ion

penetrates the silicon nitride during programming of the silicon nitride device. In an nvSRAM, nonvolatile programming only occurs during the store portion of a power down cycle. Therefore, the probability of a silicon nitride short is remote.
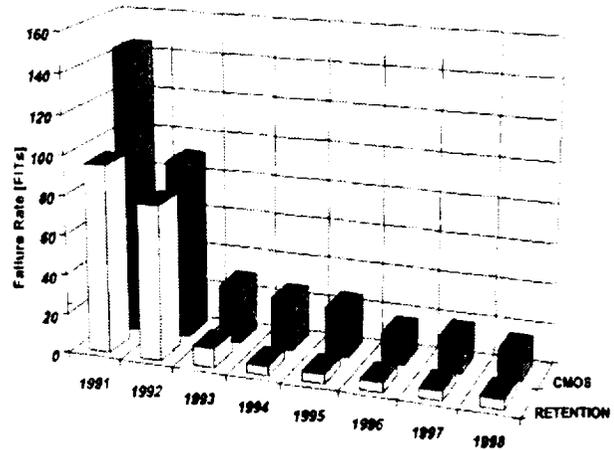
Figure 5. Failure rates of nvSRAM product supplied by Simtek.

6. CONCLUSIONS

Modern nvSRAMs have been in commercial production for approximately ten years. Radiation hardened versions of nvSRAMs can replace other forms of semiconductor non-volatile memories for earth orbit and near and deep space missions. The nvSRAMs, when combined with radiation hardened CMOS, can provide fast programming times, low power operation, virtually unlimited endurance, and resistance to total dose, proton and high energy particle irradiation. SONOS EEPROMs have been previously produced using radiation hardened semiconductor processes, and nvSRAMs are fabricated using the same semiconductor process used for SONOS EEPROMs. Only standard processing equipment typically available in semiconductor factories is needed. Therefore, the effort required to develop radiation hardened nvSRAMs is primarily in the design. New programs must be funded to develop and produce radiation hardened nvSRAMs.

7. REFERENCES

[1] P. J. McWhorter et al., "Retention Characteristics of SNOS Nonvolatile Devices in a Radiation Environment," IEEE Trans. Nucl. Sci. NS-34, 1652 (1987).

[2] G. J. Brucker, "Interaction of Nuclear Environment with MNOS Memory Device," IEEE Trans. Nucl. Sci. NS-21, 186 (1974).

Implementation of Ferroelectric Memories for Space Applications

Stephen C. Philpy, Gary F. Derbenwick, David A. Kamp and Alan F. Isaacson

Celis Semiconductor Corporation
5475 Mark Dabling Blvd., Suite 102
Colorado Springs, CO 80918
(719) 260-9133; (719) 593-8540 (fax); celis@celis-semi.com

Abstract–Ferroelectric random access semiconductor memories (FeRAMs) are an ideal nonvolatile solution for space applications. These memories have low power performance, high endurance and fast write times. By combining commercial ferroelectric memory technology with radiation hardened CMOS technology, nonvolatile semiconductor memories for space applications can be attained. Of the few radiation hardened semiconductor manufacturers, none have embraced the development of radiation hardened FeRAMs, due a limited commercial space market and funding limitations. Government funding may be necessary to assure the development of radiation hardened ferroelectric memories for space applications.

TABLE OF CONTENTS

1. INTRODUCTION

As the missions of the National Aeronautics and Space Administration (NASA) grow more complex, new on-board spacecraft computing systems are being developed. These new systems require large amounts of nonvolatile semiconductor memory that is radiation tolerant and that can operate continuously during long missions.[1] Only a few radiation tolerant nonvolatile technologies have been developed that could be useful for this purpose. SONOS EEPROMs are radiation tolerant, but cannot withstand continuous operation over extended periods due to endurance limitations. FeRAMs can potentially offer a better solution for these missions. By combining commercial ferroelectric semiconductor technology with radiation hardened CMOS technology, a more ideal nonvolatile memory for space applications can be achieved.

Since the ferroelectric storage element itself is inherently radiation tolerant, an integrated semiconductor memory can be manufactured that requires little or no external shielding for most space applications. In addition to virtually unlimited endurance, ferroelectric memories are also ideal for space because of their low power performance and fast programming times. Because ferroelectric devices have been shown to be tolerant to ionizing radiation, proton irradiation and high energy ion exposure, they are a good choice for semiconductor memories used in earth orbit as well as for near and deep space missions.

Recent advances in multi-level metal FeRAMs using SBT (strontium bismuth tantalite) as well as better electrode technology for FeRAMs using PZT (lead zirconium titanate) have been made making higher density FeRAMs a reality. The new electrode technology for PZT gives PZT high endurance, better retention, and lower voltage performance.

At present, of the few manufacturers of semiconductors for space applications, none are manufacturing memories using ferroelectric technology due to their involvement with other technologies or their concern regarding the future size of the space market. Government funding is probably necessary to get the immediate involvement of these few companies in the development of high density radiation hardened FeRAMs.

2. PERFORMANCE

NASA prefers to use a nonvolatile memory with many of the characteristics of today's Dynamic Random Access Memories (DRAMs) or Static Random Access Memories (SRAMs).[1] The performance characteristics of FeRAMs are comparable to those of DRAM's, except that FeRAMs also provide nonvolatility. Commercial FeRAMs are being developed that use one transistor, one capacitor

memory cells similar to those used in DRAMs. However, for space applications, a more rugged architecture using a pair of the memory cells is desirable. This two transistor and two capacitor (2T/2C) combination with a radiation hardened sense amplifier, shown in Fig. 1, provides for better immunity to single event upset.
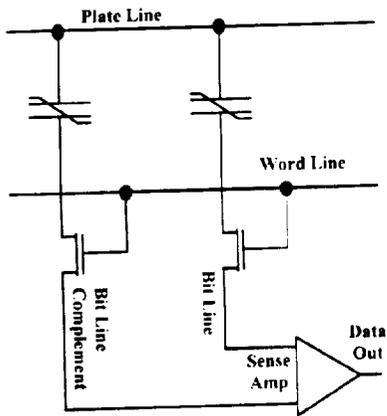

Figure 1. Schematic of a Two Transistor, Two Capacitor (2T/2C) Ferroelectric Memory Cell

Low power memories are a necessity for space applications, due to the limitation of total onboard spacecraft power. FeRAMs operate with low power because of their lower programming voltages and faster programming times, as compared to EEPROM or Flash memory.

FeRAM programming times are typically comparable to those of DRAMs. Programming times for FeRAMs designed for space applications are projected to be a few hundred nanoseconds because of the extended low temperature range that is required for space applications (-55 °C) as compared to those for commercial applications (0 °C or 20 °C).

Retention and endurance performance characteristics are key attributes for nonvolatile memories used in space applications, especially for deep space missions. Commercial FeRAMs can be fabricated with virtually unlimited endurance and retention times exceeding ten years at 125 °C.

3. RADIATION HARDNESS

All space environments include some level of radiation exposure. Depending on the length and purpose of the mission, radiation tolerant memories play a major role in the definition of the mission. Although Flash memories are the highest density nonvolatile memories being produced today, they are not tolerant to radiation exposure.[1] Ferroelectric memory cells are inherently radiation tolerant and can withstand inoizing total dose

irradiation up to or greater than 10 Mrads,[2,3] as shown in Fig. 2. Ferroelectric 1K embedded memories, in an unbiased state, have been subjected to single event upset (SEU) heavy ion exposure of an effective LET up to 128 MeV-cm^2/mg with no bit failures.[3,4] These 1K embedded FeRAM memories were also tested, in an unbiased state, to simulate solar wind, where they showed no failures due to proton irradiation up to 1.55^{12} protons/cm^2.[3,5]

Total Dose Response of SBT (Y1)
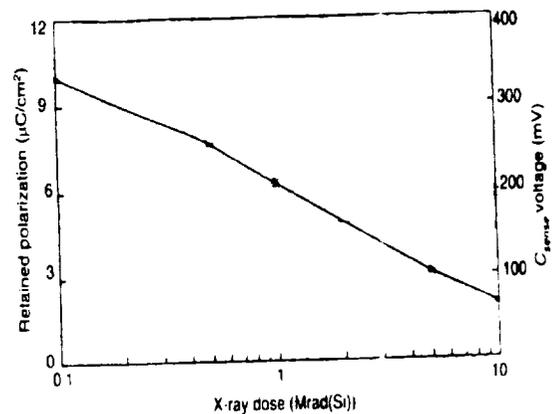
Figure 2. Degradation of SBT Capacitor Polarization Due to Radiation

The ferroelectric process module occurs after the underlying CMOS devices are fabricated but before the metalization steps. No special radiation hardened techniques are required for the ferroelectric process module.

Previously it has been demonstrated that radiation hardened CMOS underlayers can be integrated with a ferroelectric process module to obtain radiation hardened FeRAMs.[3] While the process of fabricating the ferroelectric module has a negligible effect on the radiation hardness of the underlying CMOS circuitry, certain ion implant used in the CMOS process need to be adjusted to compensate for the ferroelectric processing. The ferroelectric module typically requires annealing steps in oxygen at temperatures greater than 650 °C. The effects of these anneals on the n-channel threshold voltage and SEU resistors are shown in Figs. 3 and 4, respectively.[3] A shift of approximately 100 mV is observed on the n-channel threshold voltages and a shift of approximately one order of magnitude is observed on the SEU resistors. Modifying the respective implant doses shifts these parameters back to their nominal CMOS values.
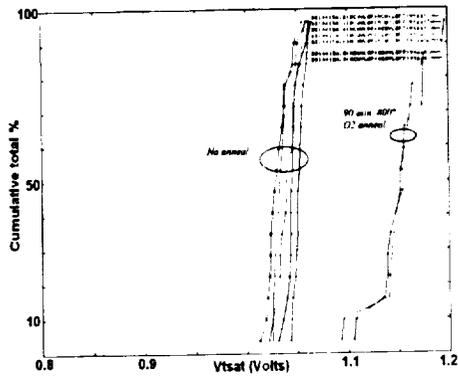
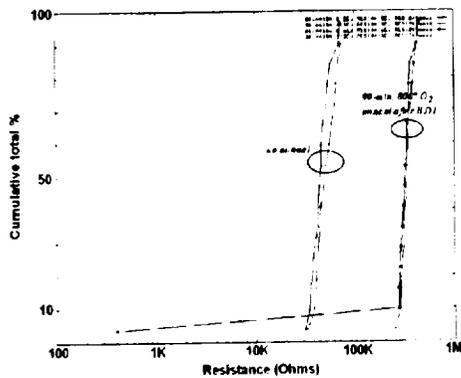Figure 3   N-Channel Threshold Voltage Shift Due to
Annealing

Figure 4   SEU Resistance Shift Due to Annealing

The ferroelectric process requires use of silicon nitride for
final chip passivation. Previously, many radiation hard-
ened CMOS processes have used silicon oxide passi-
vation. Fig. 5 shows that there is no degradation of the
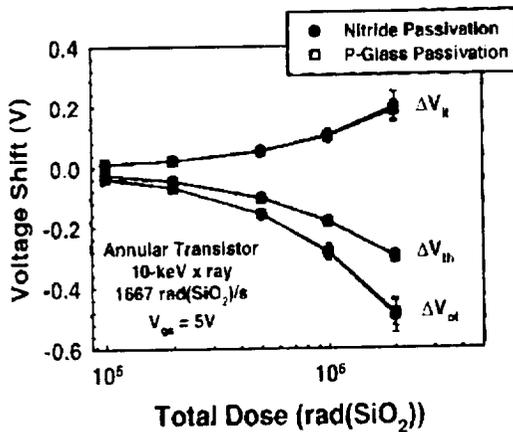radiation hardness due to the use of silicon nitride
passivation.[3]

Figure 5   No Degradation Due to Nitride Passivation
of SBT Process

4. SPACE COMMERCIALIZATION

Today, many major companies have announced FeRAM
commercialization programs worldwide. These activities
provide a critical mass of manpower and expenditures to
advance FeRAM development for commercial pro-
duction. Although there are other novel nonvolatile
memory technology development efforts being worked on
that may have application for space environments, today
these other efforts lack the maturity that has been gained
regarding the development of FeRAMs

Commercial FeRAM developers are not currently
working on radiation tolerant versions. These semi-
conductor manufacturers are not traditionally concerned
with radiation effects, as is the case with most off shore
companies. Currently there are only two commercial
semiconductor manufacturers located in the United States
that can produce radiation tolerant memories to greater
than 1 Mrad of ionizing dose.

Previous work involved the development of a radiation
hardened 128K FeRAM, using a split processing
method.[3] The radiation tolerant CMOS underlayers were
produced in a United States semiconductor facility and a
commercial offshore ferroelectric wafer fabrication
facility would finish the fabrication of the memories using
SBT (strontium bismuth tantalite) ferroelectric technology
and single level metal interconnect technology. This
previous work was terminated early due to program
funding changes.

Since some of the new spacecraft computing systems
being developed today require larger densities of radiation
tolerant nonvolatile memory than can be provided by a
128K FeRAM, future development efforts for radiation
hardened FeRAMs should target multi-Megabit densities
using multi-level metalization. Successful commercial
FeRAMs using multi-level metalization have recently
been reported[6].

5. CONCLUSIONS

FeRAM memories have performance characteristics that
are desirable for space applications, including low power
operation, virtually unlimited endurance and fast
programming times. When combined with radiation
hardened CMOS, FeRAMs meet stringent space radiation
requirements. Insertion of the ferroelectric processing
module into a radiation hardened CMOS process causes
only minor perturbations of the CMOS process that can
be changed by simple changes in a few ion implant doses.
Either SBT or PZT technologies can be used for the
ferroelectric storage element in radiation hardened
FeRAMs.

Recent advances in manufacturing commercial FeRAMs have included multi-level metalization required for higher density radiation hardened FeRAMs. Government funding is probably needed to develop radiation hardened FeRAMs at a United States memory manufacturer that has radiation hardened CMOS processing capability.

6. ACKNOWLEDGMENTS

7. REFERENCES

[1]K. F. Strauss and T. Daud, "Overview of Radiation Tolerant Unlimited Write Cycle Non-Volatile Memory," *2000 IEEE Aerospace Conference* (2000).

[2]SBT cell data courtesy of J. M. Benedetto and UTMC. (1998)

[3]S. C. Philpy, D. A. Kamp, A. D. DeVilbiss, A. F. Isaacson and G. F. Derbenwick, "Ferroelectric Memory Technology for Aerospace Applications," *2000 IEEE Aerospace Conference* (2000).

[4]J. M. Benedetto, G. F. Derbenwick and J. D. Cuchiaro, "Single Event Upset Immunity of Strontium Bismuth Tantalite Ferroelectric Memories," *IEEE Transactions on Nuclear Science* (1999).

[5]The authors would like to thank Gary Swift of the Jet Propulsion Laboratory for his collaboration regarding these experiments (1999).

[6]S. S. Lee and N. S. Kang, FeRAM Team, Memory R&D Division, Hyundai Electronics Industries Co., Ltd. and S. C. Philpy and G. F. Derbenwick, Celis Semiconductor Corporation, "Reliability and Performance of a 64K Ferroelectric Memory" *JPL 2000 Micro-electronics Reliability & Qualification (MRQ) Workshop* (2000).

SESSION #3: *Innovative Concepts*

Wednesday, November 15, 2000
3:40 PM – 6:00 PM
Chair: *Lynn Adams*, SEAKR Engineering,
Englewood, Colorado

3:40 PM	Characterization of a New EEPROM/Flash Memory Cell; *R. Reedy, J. Cable, F. Wright, H. Anthony and C. Tabbert*, Peregrine Semiconductor
4:00 PM	SOI Non-volatile Memory; *P. McMarr, R. Lawrence, H. Hughes and W. Jenkins*, Naval Research Laboratory
4:20 PM	Mechanisms of Protonic Nonvolatile Memory Device; *P. Macfarlane and R. Stahlbush*, Naval Research Laboratory
4:40 PM	Functional Nanostructured Particle Arrays for Nanocrystal Nonvolatile Memories; *H. Atwater*, California Institute of Technology
5:00 PM	Quantitative Analysis of Charge Injection and Discharging of Si Nanocrystals and Arrays by Electrostatic Force Microscopy; *L.D. Bell*, Jet Propulsion Laboratory, *E. Boer, M. Ostraat, M. Brongersma, R. Flagan and H. Atwater*, California Institute of Technology
5:20 PM	Advanced Compact Holographic Data Storage System; *T.H. Chao, H. Zhou and G. Reyes*, Jet Propulsion Laboratory
5:40 PM	Ultra-High Density Holographic Memory Module in Solid-State Architecture; *V. Markov*, MetroLaser

Characterization of a New EEPROM/Flash Memory Cell

Ron Reedy, Jim Cable, Frank Wright, Hal Anthony & Chuck Tabbert
Peregrine Semiconductor Corporation
6175 Nancy Ridge Drive, San Diego, CA 92121

Abstract

This paper discusses on-going work sponsored through the Defense Threat Reduction Agency (DTRA) SBIR program at Peregrine Semiconductor Corp to characterize the radiation and endurance performance characteristics of Peregrine's patented *PlusCell* ^{TM} - a new non-volatile memory cell which utilizes a standard CMOS process on a silicon on insulator technology.

This work begins with basic cell characterization of write voltages & read currents on sample cells across multiple wafer lots, progresses through cell retention and endurance testing across temperature and radiation environment, extracting necessary activation energy for cell charge leakage and completes with total dose characterization of the basic cell through 100K write/erase cycles with radiation testing performed after 0, 10K & 100K write/erase cycles.

The *PlusCell* ^{TM} is very compact and operates at write voltages as low as 6 V and read voltages below 1 V. Charge injection of both polarities is accomplished through a 100 A thick gate oxide by means of hot carrier injection. The cell relies on isolation provided by an insulating substrate. In the case of Peregrine's UTSi® technology, the base technology is fully depleted silicon on sapphire CMOS and the cell is manufactured without any additional processing or masking steps. Measured cell performance of charge retention, radiation sensitivity and endurance is presented.

Traditional EEPROM or flash cells require specialized regions of thin tunneling oxide, which increase complexity and cost while reducing density and yield. Additionally, the normal injection mechanism is based only on electrons, which means writing and erasing must be carefully balanced, or over-erasing can occur. The *PlusCell*^{TM}, using both hole and electron injection, has several key advantages including elimination of any over-erase mechanism, an extremely dense cell, availability of bi-directional read, and efficient block erase.

Use of EEPROM at Peregrine Semiconductor

It is intended that this cell will be basis for embedded EEPROM applications by the Space & Defense community along with introductions of high-density EEPROM devices by Peregrine Semiconductor Corp.

DTRA Program Plan & Results To Date

The work plan has been divided in specific tasks detailed below:

Task 1 – Basic Cell Characterization

Subtask 1.1
Objective: Characterize write voltages and read currents on a sample EEPROM cell across wafer lots.

Characterization of an Intrinsic N-Type Cell was chosen for this subtask with associated external connection schemes (Figures 1 & 2)

enable the cell. Moreover, there is a high level of repeatability across wafer and lots. The following histograms shows read currents as a function of write voltage for both P & N channel devices
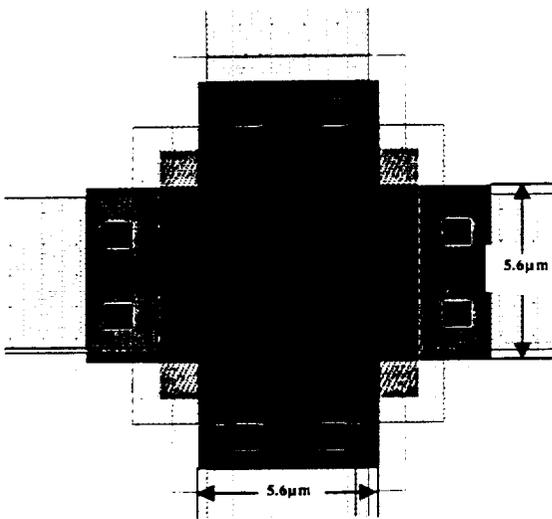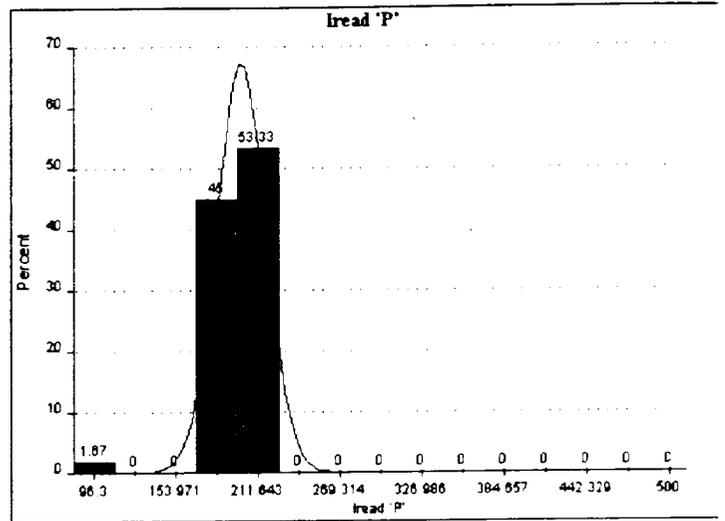

Figure 1. Selected IN5x5μm E^2 Cell for DTRA Characterization
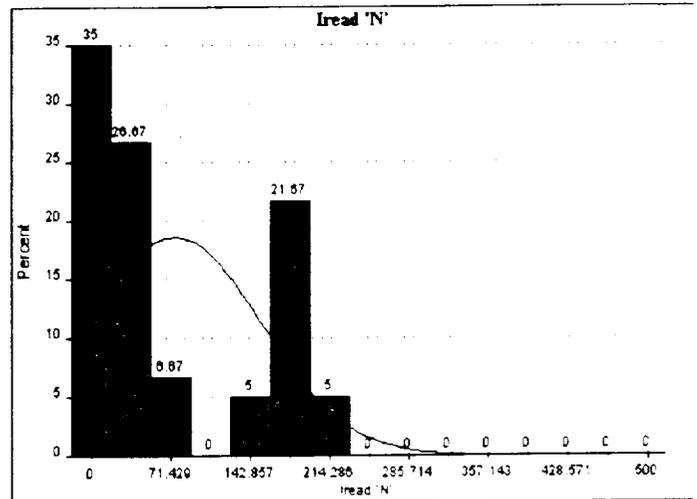

Figure 3 – P Channel 9 Volt Write Pulse

Figure 2. Test Fixture connections to E^2 cell.

Data was taken across a variety of samples and a variety of wafer lots and the following data concluded that when the *PlusCell^{TM}* is in a known state (N and P are low), there is a high level of confidence that a 15mS write pulse with amplitudes of 10 and 9 volts respectively will

Figure 4 – N Channel 10 Volt Write Pulse

66

Subtask 1.2

Objective: Characterize retention and endurance on a sample cell, including extraction of activation energy for cell charge leakage. Cell retention will be measured at three temperatures (85C, 125C, and 150C) following 0, 10K, and 100K write/erase cycles. Retention will be measured out to 168 hours.

Sample cells data was take across temperature. Since the time to fail point had been arbitrarily determined to be 20% reduction in read current and the initial read current of each test cell varies, the data must be normalized to represent an equal starting point for each cell at a given high test temperature. The following two semi-log plots show the normalized data of the P and N channels respectively.

Table 1: P-Channel Data Retention

Table 2 N-Channel Data Retention

Determination of E_A:

The normalized semi-log plots give the equations for the trend lines and the R^2 values.

The equation for the P channel trend line at T=85°C is:

$$y = -0.0139Ln(x) + 0.9641$$

By arbitrarily choosing a 20% reduction in read current as the failure condition, we get

$$y = 0.8$$

Substituting y = 0.8 and solving equation 1 for x, Time To Fail values are determined:

$$x = 138.9 \times 10^3 \text{ Hrs. (T = 85°C Curve)}.$$

Similarly solving the same form of equation 1 for the 150°C curve:

$$x = 225 \text{ Hrs.}$$

Since temperature is the only failure mechanism, a simple form of the Arrhenius Equation can be used to determine E_A.

$$AF = \frac{TTF_{85}}{TTF_{150}} = e\left(\frac{E_A}{k}\left(\frac{1}{T_{85}} - \frac{1}{T_{150}}\right)\right)$$

(2)

where: AF = Acceleration factor

k = 8.617 eV/K (Boltzmans Constant)

T = Kelvin

Calculating AF and solving equation 2 for E_A, we find:

E_A = 1.29eV for the P channel device

Similarly:

E_A = 0.71eV for the N channel device

The Arrhenius Plot can be used to determine the Activation Energy as well. The gradient of the line between the data points yields the E_A value in eV. Obviously additional data samples would yield a more precise value of E_A, however this data get us in the ballpark. The following two Arrhenius plots yield E_A values that correlate well with the calculated values. Calculated values: N = 0.71eV, P = 1.29eV
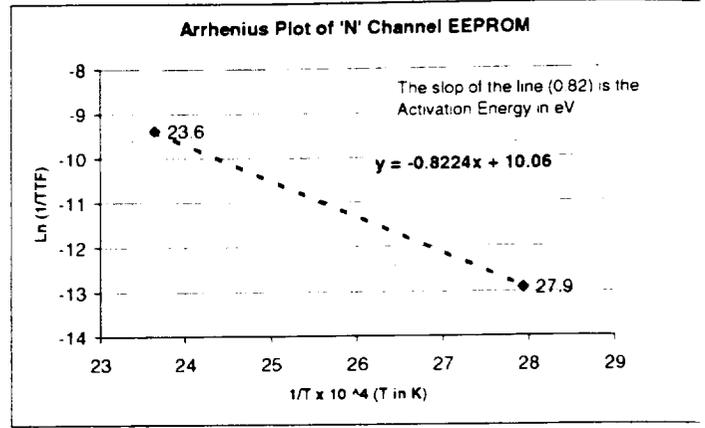
Table 3 – Arrhenius Plot of 'N' Channel EEPROM
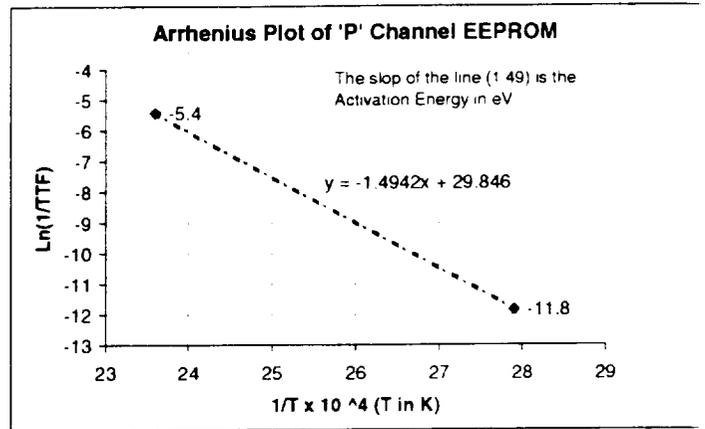
Table 4– Arrhenius Plot of 'P' Channel EEPROM

Task 2   Radiation Testing

Subtask 2.1

Objective: Characterize effect of total dose irradiation on basic cells. Radiation testing to be performed after 0, 10K, and 100K write/erase cycles. Radiation testing to be performed to 100Krad, 300Krad, and 1 Mrad. Following irradiation, cell retention will be measured at 125C for up to 168 hours. Both X-ray and Co-60 testing will be performed.

Work to be reported on in later paper

68

<u>Subtask 2.2</u>

Objective: Characterize radiation response of charge pump circuits. Measure charge pump short circuit current and open circuit voltage at total dose irradiation levels of 0, 100Krad, 300Krad, and 1 Mrad.

Work to be reported on in later paper.

<u>Conclusions</u>

Initial characterization data shows that the *PlusCell*^{TM} should become a viable non- volatile memory cell for use in military and space applications. Further characterization will define robustness and reliability of cell and follow-on circuit applications.

SOI Non-volatile Memory

P. J. McMarr, R. K. Lawrence, H. L. Hughes and W. C. Jenkins

Naval Research Laboratory, Washington, DC 20375-5000

hughes@estd.nrl.navy.mil   202-767-2429

Abstract—SIMOX silicon-on-insulator (SOI) substrates were implanted with silicon ions at doses of 5×10^{15} and 1×10^{16} ions/cm² at 130 keV. At this energy, the majority of the silicon ions were implanted into the buried oxide of the SIMOX substrate. Silicon-oxide-silicon test structures were fabricated on these implanted substrates. Point-contact transistor (PCT) measurements were performed on the test structures and initial threshold voltages determined. A stress voltage was then applied to the test structure for times ranging from 10 millisec to 100 sec. After removal of the stress voltage, PCT measurements were again performed. The threshold voltages shifted by significant and reproducible values. This stress-voltage-induced device switching response forms the basis for a new type of non-volatile SOI memory. The temperature stability of this switching was studied from –50 °C to +200 °C. In addition, the field and time dependence of the switching is also presented.

1. INTRODUCTION

MOS devices fabricated on SOI substrates have speed and power advantages compared to devices fabricated on bulk silicon. In addition, in a space radiation environment MOS devices fabricated on SOI substrates show enhanced immunity to single event effects compared to devices fabricated on bulk silicon.[1]

Bulk design rules have been ported to SOI substrates with clever modifications implemented to take into consideration effects that are inherent to partially depleted SOI, such as floating body effects.[2] Unique devices, that use the dielectric isolation of SOI substrates, have been fabricated that are difficult or impossible to build in bulk.[3]

In this study, the authors use the buried oxide of SIMOX SOI substrates to form a non-volatile memory. An experimental test device is first described which is used to demonstrate the non-volatile memory. The use of this device is advantageous with respect to the high fabrication cost of MOS device lots. Next, the modifications to the SIMOX substrate that produce a switching characteristic in the experimental test device are described. The field dependence of the switching is shown and the temperature stability of the switching is

studied over the available operable range. Finally, the switching speed characteristics, within present instrumentation limits, are defined.

2. SOI POINT-CONTACT TRANSISTOR

The purpose of this study is to build a new type of non-volatile memory using SIMOX substrates. Details relating to the manufacture of SIMOX substrates, SOI device performance advantages, and other topics can be found in the literature.[4]

The test device used in this study was invented by Liu et al.[5-8] and is called the SOI point-contact transistor. A diagram of the point-contact transistor is shown in Fig. 1.
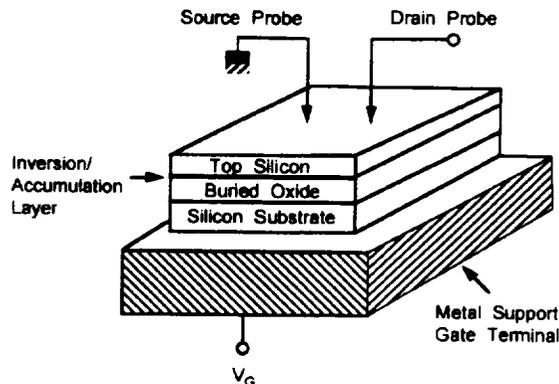

Figure 1. Schematic of the point-contact transistor.

The MOS structure inherent in all SOI substrates is used with the bulk silicon substrate as the gate terminal, the buried oxide as the insulator, and the top silicon as the transistor body. To perform point-contact transistor measurements on a sample from a wafer, the top silicon must be isolated from the edge of the sample. Leakage paths exist around the edge of the top silicon and these short-circuit the point-contact transistor. In addition, to easily extract process dependent parameters from point-contact transistor measurements, such as interface state density at the back-channel, the top silicon should be fully

depleted. This puts limits on the thickness and doping level of the top silicon.[5,8]

In this study, the SIMOX wafers were p-type (2-6 Ωcm) (100) silicon. The top silicon layer was ~1500 Å thick and the buried oxide thickness was ~1700 Å. Samples (1 cm^2) were cut from the center, the top, bottom and sides of a wafer. To isolate the top silicon on these samples, aluminum was first deposited on the top silicon as circular pads with an area of 0.005 cm^2. The patterned samples were placed in hydrazine, a highly specific silicon etch. The aluminum was then removed with hot phosphoric acid. This procedure formed the silicon-insulator(oxide)-silicon (SIS) test structures.

The thickness and the doping level of the silicon of the SIMOX wafers were within the levels necessary to assure that the top silicon was fully depleted and the SIS fabrication procedure isolated the top silicon from the edges of the sample. A typical I_D-V_G curve from point-contact transistor measurements performed on a SIS structure is shown in Fig. 2.
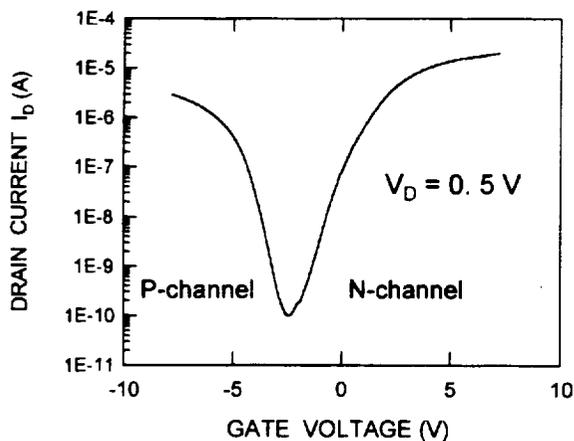
Figure 2. Subthreshold characteristics of a point-contact transistor from measurements performed on a SIS structure fabricated on the SIMOX substrate (applied drain voltage (V_D) = 0.5 V).

As shown in Fig.2, depending upon the gate bias, accumulation or inversion channels are activated at the top-silicon/buried-oxide interface (the back-channel). The variation of current with gate voltage is exponential and the density of back-channel interface states can be determined from the subthreshold swing. Using the results in Fig. 2, this value is < 0.6 V per decade, which corresponds to a back-channel interface state density of ~ 10^{11} $cm^{-2}eV^{-1}$. The small leakage current (minimum value in Fig. 2 is ~10^{-10} amps), and the fact that the minimum of the current occurs near zero gate voltage, further confirm that the top silicon is fully depleted.[9] In addition, the drain current minimum near zero gate voltage indicates there is no detectable net trapped charge in the buried oxide.

Figure 3 is a plot of drain current versus gate voltage, in strong inversion and accumulation, as a function of drain voltage. The intercepts with the V_G axis correspond to the threshold voltage V_T for the inversion channel and the flat-band voltage V_{FB} for the accumulation channel.[7,8] From Fig. 3, V_T ~ - 4.5 V and V_{FB} ~ + 2.0 V.
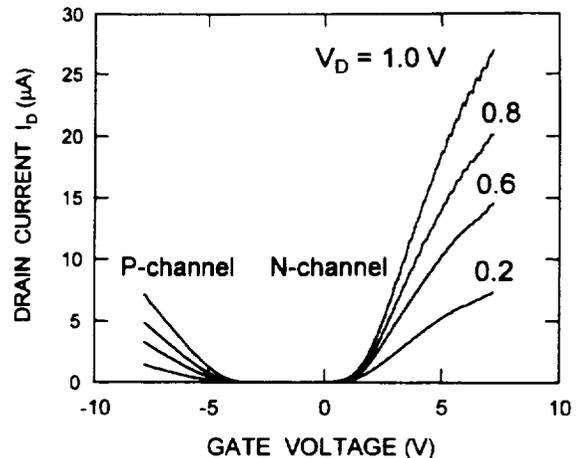
Figure 3. Drain current (I_D) versus gate voltage (V_G) as a function of drain voltage (V_D) from point-contact transistor measurements.

3. MODIFIED SIMOX SUBSTRATES

Thermal SiO_2 implanted at a high dose exhibits hysteresis in I-V and C-V characteristics and has been studied for possible application as non-volatile memory.[10] The hysteretic behavior in the Si^+ thermal oxides was induced by the application of a stress voltage for various times. This phenomenon is well documented,[11,12] and was the impetus for the following sequence of experiments.

Following the initial electrical characterization of the SIMOX substrates, the substrates were implanted with silicon ions at doses of 5×10^{15} and 1×10^{16} ions/cm^2 at 130 keV. The SIMOX substrates selected for implantation had a top silicon thickness of ~1500 Å and a buried oxide thickness of ~1700 Å. For this system, at 130 keV, TRIM and SILVACO simulations showed that the majority of the silicon ions would be implanted into the buried oxide. SIMS analysis confirmed that ~90 % of the ions were implanted in the top 600 Å of the buried oxide. The temperature of the SIMOX substrates during implantation was 200°C. The ions were channeled to minimize damage in the top silicon of the substrate.

To determine if the electrical properties of the substrates were degraded by the high dose implantations, SIS structures were fabricated on the implanted substrates and PCT transistor measurements performed. A typical result is shown in Fig. 4. There is a small change in the subthreshold swing, which corresponds to a back-channel

interface state density of $\sim 3 \times 10^{11} cm^{-2} eV^{-1}$. The magnitudes of the currents above the subthreshold region are essentially identical for the implanted and unimplanted substrates, indicating that the implantation process does not seriously impact carrier mobility.
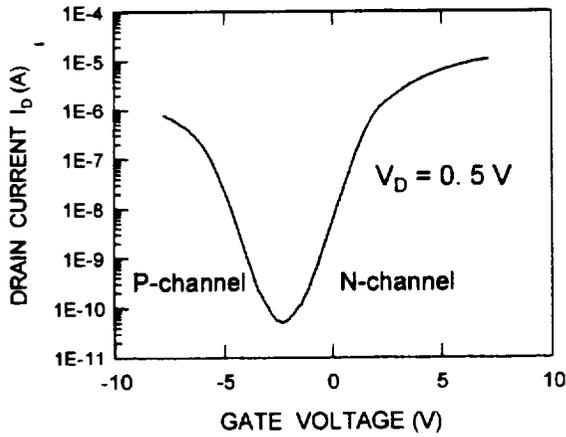

Figure 4. Subthreshold characteristics of a point-contact transistor from measurements performed on a SIS structure fabricated on a Si^+ implanted SIMOX substrate (dose = $1 \times 10^{16} / cm^2$).

4. SWITCHING CHARACTERISTICS OF THE IMPLANTED SIMOX SUBSTRATES

Following initial electrical characterization of the implanted substrates, a stress voltage of +25 V was applied to the back-gate (BG) of the PCT for 10 sec. A PCT measurement was then performed, a stress voltage of -25 V was then applied to the back-gate and another PCT measurement performed. The results are shown in Fig. 5.

When the BG is biased at +25 V, the I_D-V_G curve is displaced from the initial unstressed value to positive gate voltage values by ~4 V. When the BG is biased to -25 V, the I_D-V_G curve is displaced to negative gate voltage values by ~4V. When the BG is biased at ± 25 V, the I_D-V_G curves show a relative displacement of ~8V. This stress-voltage-induced switching is reproducible, as shown in Fig 6.

The field dependence of the switching for positive stress voltages is shown in Fig. 7. PCT measurements were performed on an unstressed SIS structure and the initial threshold voltage determined. A positive bias was then applied to the BG for 10 sec. The bias was removed, PCT measurements were performed, and the change in the threshold voltage determined. The bias range selected was +5 V to +30 V. Figure 7 shows that for a stress voltage of +5 V there is no change in the threshold voltage (no switching) and that for a stress voltage of +30 V the threshold voltage difference has increased to ~6 V.
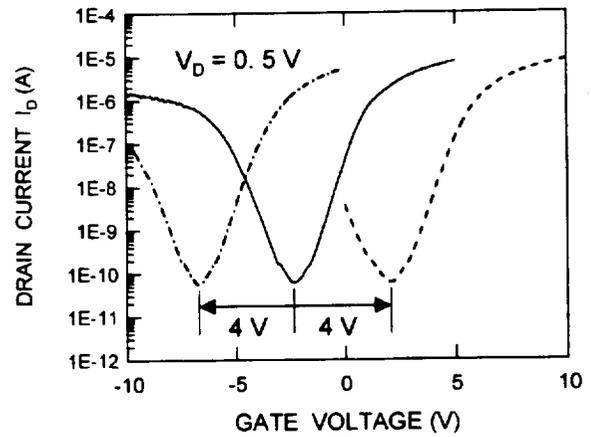
Figure 5. Subthreshold characteristics for an unstressed implanted substrate (solid curve), +25 V stress (dashed curve), and -25 V stress (dashed-dotted).
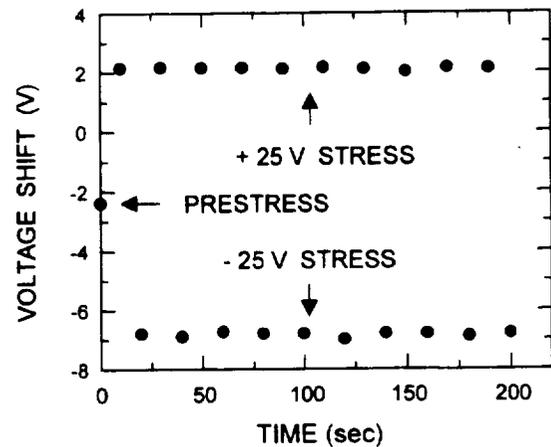
Figure 6. Switching characteristics of an implanted substrate stressed at ± 25 V for 10 sec (room temperature).
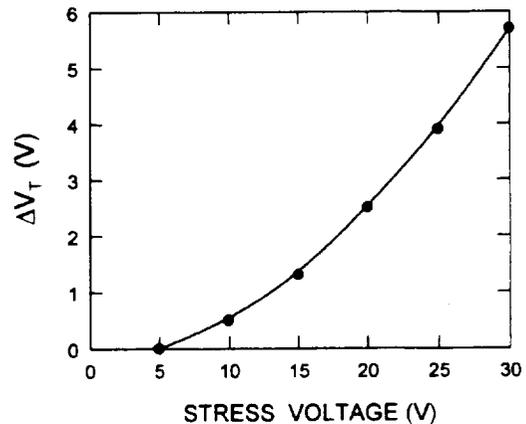
Figure 7. Switching characteristics of an implanted substrate for stress voltages from +5 V to +30 V (10 sec).

Fringing fields exist between the source and drain and the channel region of MOS devices fabricated on SIMOX or other SOI substrates.[4] These fringing fields extend from the source and drain through the buried oxide, intersect the top-silicon/buried-oxide interface beneath the channel region, and terminate in the channel region. Any charge at the top-silicon/buried-oxide interface beneath the channel region, therefore, has an influence on the threshold voltage of devices fabricated on SOI substrates.[13]

Figure 7 shows that a stress voltage of 5 V (or less) has no affect on the threshold voltage of the SIS devices. This means that there is no detectable charge at the at the top-silicon/buried-oxide interface. Thus, partially or fully depleted devices, operating at a V_{DD} of 5 V or less, could be fabricated on the Si^+ implanted substrates with no performance degradation.

5. TEMPERATURE DEPENDENCE

To determine the temperature stability of the switching, PCT measurements were first performed on an unstressed Si^+ substrate from –50 °C to +200 °C. Figure 8 shows the threshold voltage shift over this temperature range for the implanted and unimplanted substrates. The threshold voltage for the P-channel decreases for both substrates as the temperature increases, with small differences between the two substrates at higher temperatures.
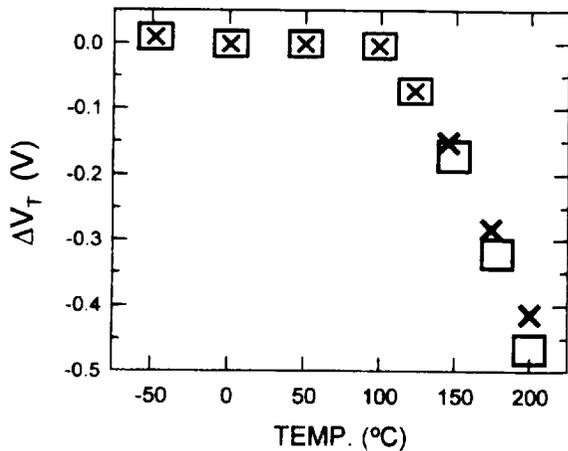
Figure 8. Threshold voltage shift as a function of temperature for an unimplanted substrate (cross) and an implanted unstressed substrate (hollow box).

The temperature dependence of the switching was now studied from –50 °C to 200 °C. The procedure followed was identical with that discussed in the preceding section. The results at 200 °C are shown in Fig. 9. These results should be compared to those in Fig. 6, where the switching characteristic of the implanted substrate was studied at room temperature. Although there is noticeable

jitter in the high temperature switching, the "safe" switching window is still large, ~6V, as compared to ~8 V at room temperature.
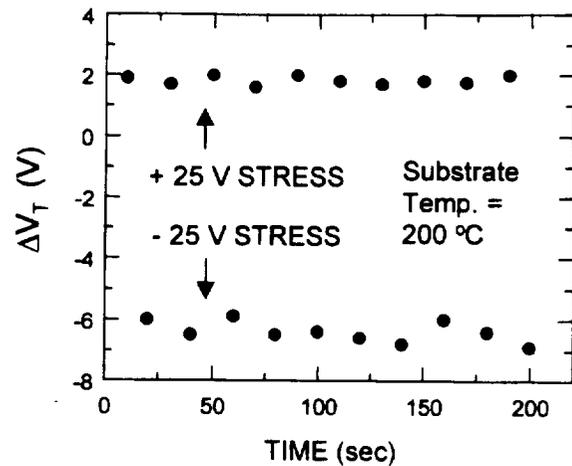
Figure 9. Switching characteristics of an implanted substrate stressed at ± 25 V for 10 sec (substrate temperature 200 °C).

6. SWITCHING SPEED

The switching speed of the SOI non-volatile memory is shown over four decades of time in Fig. 10. The switching window decreases from ~8 V when the stress voltage is applied for 10 sec, to ~ 1.4 V when the stress voltage is applied for 10 millisec. Voltage stressing at ± 25 V for shorter time periods is not possible with available instrumentation.
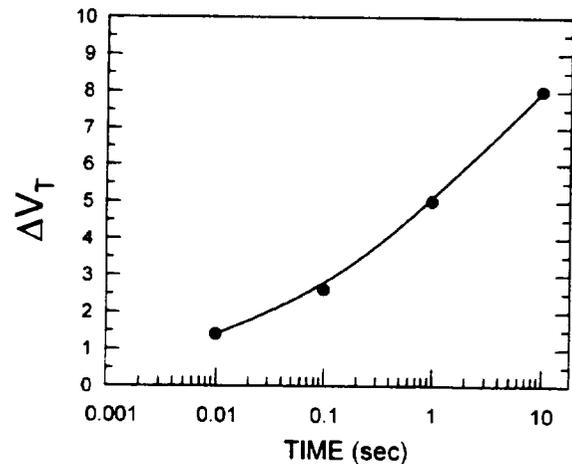

Figure 10. Threshold voltage shift as a function of the time the stress voltage was applied (stress voltage = ± 25 V).

Although the window at 10 millisec has decreased to ~1.4 V, the threshold voltage variations of MOS devices are much more tightly controlled than this, and these substrates could function as non-volatile memory (with minimum verified switching speed of 10 millisec).

7. DISCUSSION AND CONCLUSIONS

With the range of available non-volatile memory technologies, the logical question to ask is "What advantages does this SOI non-volatile memory afford beyond the established technologies?"

The first advantage is the use of state-of-the-art SOI substrates. Devices fabricated on SOI substrates are faster, consume less power, and ultimately may prove cheaper than bulk silicon.

The second advantage is the robustness of the manufacturing procedure. There are several different technologies that rely upon silicon implantation to form non-volatile memory. Some, as discussed in the text, implant silicon into the gate oxide[10] Others use silicon nano-crystals as memory units.[14,15] Both procedures rely upon tunneling from the silicon substrate through an oxide. Although successful, this necessarily puts tight constraints on fabrication procedures. The implantation must be carefully controlled so as not to damage the oxide and the oxide/silicon interface. The thermal oxides are thick, and although the units can be integrated with MOS processing, this is an additional, costly step that may not be competitive with conventional flash memory.

In the present implantation procedure, damage to the top silicon is the main consideration. As discussed in the text, this can be minimized by channeling, something that, by definition, cannot be accomplished with a noncrystalline oxide. In addition, due to the confined nature of the buried oxide, any damage to the top-silicon/buried-oxide interface can be easily annealed out, with no subsequent degradation of the buried oxide.

An additional, subtle advantage of this SOI nonvolatile memory was briefly discussed at the end of section 4. In that section the point was made that using the implantation parameters discussed in section 3, both partially depleted and fully depleted devices could be fabricated on the same substrate as the non-volatile SOI memory.

An analogue of the PCT can be fabricated by simply selectively removing the gate oxide on a FET and metalizing the area that was covered by the gate oxide. This would serve as the terminal where the stress voltage is applied. When the stress voltage is removed, the source and drain of the FET would then function just as the metal probes do in the PCT. A smaller voltage would then be applied to the metalized region to create a channel at the top-silicon/buried-oxide interface. Although this type of device has not been fabricated, the concept seems relatively straightforward.

The advantage this gives is that the non-volatile memory can be "on demand." Every FET has the potential to be a storage device. This integrates the non-volatile memory directly with other devices using established silicon processing technology.

8. ACKNOWLEDGMENTS

9. REFERENCES

1. A. J. Auberton-Herve, "SIMOX-SOI Technologies For High Speed and Radiation Hard Technologies: Status and Trends in VLSI and ULSI Applications," *Proceedings of the 4th International Symposium on SOI Technology and Devices, Electrochemical Society*, Vol. 90, No. 6, pp. 455-478, May 1990.

2. M. R. Casu, G. Masera, G. Piccinini, M. Ruo Roch, and M. Zamboni, "Comparative Analysis of PD-SOI Active Body-Biasing Circuits," *Proceedings of the 2000 IEEE International SOI Conference*, pp. 94-95, October 2000.

3. X. Huang, W-C. Lee, C. Kuo, D. Hisamoto, L. Chang, J. Kedzierski, E. Anderson, H. Takeuchi, Y-K. Choi, K. Asano, V. Subramanian, T-J. King, J.Bokor, and C. Hu, "Sub 50-nm FinFET: PMOS," IEDM Technical Digest, pp. 67-70, December 1999.

4. Jean-Pierre Colinge, *SILICON-ON-INSULATOR TECHNOLOGY: Materials to VLSI, 2nd Edition*, Kluwer Academic Publishers, Boston/ Dordrecht/ London, 1997.

5. S. T. Liu, P. S. Fechner, and R. L. Roisen, "Fast Turn Characterization of SIMOX Wafers," *Proceedings of the 1990 IEEE International SOS/SOI Technology Conference*, pp. 61-62, September 1990.

6. S. T. Liu, J. C. Yue, and P. S. Fechner, Patent US5519336, "Method for Electrically Characterizing the Insulator in SOI Devices," issued May 21, 1996.

7. S. Cristoloveanu and S. Williams, "Point-Contact Pseudo-MOSFET for In-Situ Characterization of As-grown Silicon-on-Insulator Wafers," *IEEE Electron Device Letters*, Vol. 13, No. 2, pp. 102-104, February 1992.

8. S. Williams and S. Cristoloveanu, "Point Contact Pseudo-Metal/Oxide/Semiconductor Transistor in As-Grown Silicon on Insulator Wafers," *Materials Science and Engineering*, Vol. B12, pp. 191-194, June 1992.

9. J. Yue, H. K. Liou, and S. T. Liu, "Screening SIMOX Wafers' Background Doping for SOI CMOS SRAM Production," *Proceedings of the 8th International Symposium on SOI Technology and Devices*, Vol. 97-23, The Electrochemical Society Series, Pennington, pp. 137-142, 1997.

10. T. Ohzone, T. Matsuda, and T. Hori, "Erase/Write Cycle Tests of n-MOSFET's With Si-implanted Gate- SiO₂," *IEEE Transactions on Electron Devices*, Vol. 43, No. 9, pp. 1374-1381, September 1996.

11. A. Kalnitsky, A. R. Boothroyd, and J. P. Ellul, "A Model of Charge Transport in Thermal SiO₂ Implanted with Si," *IEDM Technical Digest*, 1988, pp. 516-519.

12. M. Hao, H. Hwang, and J. C. Lee, "Memory Effects of Silicon-implanted Oxide For Electrically Erasable Programmable Read-only Memory applications," *Applied Physics Letters*, Vol. 62, No. 13, pp. 1530-1532, August 1993.

13. J. H. Smith, R. K. Lawrence, and G. J. Campisi, "Numerical Analysis of Silicon-On-Insulator Short Channel Effects in a Radiation Environment," *Journal of Electronic Materials*, Vol. 21, No. 7, pp. 683-687, May 1992.

14. K. Han, I. Kim, and H Shin, "Programming Characteristics of P-channel Si Nano-crystal Memory," *IEEE Electron Device Letters*, Vol. 21, No. 6, pp. 313-315, June 2000.

15. S. Tiwari, F. Rana, K. Chan, and O. Buchanan, "Volatile and Nonvolatile Memories in Silicon With Nano-crystal Storage," *IEDM Technical Digest*, pp. 521-524, 1995.

Mechanisms of Protonic Nonvolatile Memory Device

P.J. Macfarlane[a] and R.E. Stahlbush, Naval Research Laboratory.
Code 6816, 4555 Overlook Ave, Washington, DC 20375
[a]pmacfarl@estd.nrl.navy.com

Abstract— A nonvolatile memory device based on protonic transport in oxides has been proposed.[1] The mobile H^+ ions are introduced into the SiO_2 layer by annealing $Si/SiO_2/Si$ structures in H_2 at temperatures greater than 500°C. This effect has only been observed for confined oxides that have been annealed at ≥ 1100°C prior to the hydrogenation anneal. This includes buried oxides such as Unibond and SIMOX as well as thermal oxides annealed with a polysilicon cap. An applied field moves the charge within the oxide and the charge stops moving when the field is removed. In a memory device, the hydrogen-annealed oxide is the gate oxide and the position of the mobile charge is sensed by the shift of the I-V curve. Much is still not understood about the motion of the charge across the buried oxide. Previous work has assumed that H^+ transport and the time it takes to traverse the oxide is governed by interactions within the bulk of the oxide. Based on parameters that affect the transport time, we conclude that H^+ trapping and detrapping at the Si/SiO_2 interface are more important than H^+ interactions within the oxide bulk. These parameters include the applied field, the H^+ concentration and the oxide thickness. One consequence is that projections of device write-time based on the previous assumptions of H^+ transport mechanisms may be overly optimistic.

TABLE OF CONTENTS

1. INTRODUCTION

Recently, a non-volatile field effect transistor (NVFETs) memory based on positive charge transport in an $Si/SiO_2/Si$ structure has been proposed.[1] The mobile positive charge, introduced into the NVFET during annealing in a H_2 containing ambient at temperatures greater than 500°C, can be transported from one Si/SiO_2 interface to the other Si/SiO_2 interface by changing the orientation of the field across the gate. Moving the positive charge in the oxide causes the threshold voltage of the device to shift, which in turn affects the drain current, increasing or decreasing it. The changes in the drain current can then be used to read the bit state to which the device had been written before the power applied to the device was removed. The NVFET has several advantages over more traditional forms of nonvolatile memory. The programming voltage required by the NVFETs is approximately a factor of 5 less than that required by EEPROM. Additionally, the hydrogen anneal used to create these new devices is similar to standard Si device fabrication processes. Retention tests of the NVFETs indicated that the devices were able to retain their bit state as either "1" or "0" after the devices were heated up to 200° C for up to 25 h.[2] The lengthy endurance of these devices suggests that there is little loss in the concentration of mobile charge during extended operation.

Previous studies of the positive charge suggest that the ions are H^+ bonded to bridging O atoms.[1] Key to creating the H^+ ions appears to be a high temperature anneal (≥ 1100°C), which the SOI material receives as part of its standard processing. This high temperature anneal is conducted before the H_2 annealing which introduces the mobile charge. These studies suggest that transport of the H^+ is dominated by its drift within the oxide which is modeled to occur by a series of "hops" as the proton moves from one O atom to another nearby O atom.[3] Protonic drift was used by McLean to account for the time dependence of interface state build-up observed after pulsed irradiation of thermal SiO_2 layers.[4] This work was later verified by the oxide thickness and field dependence studies of Saks and Brown.[5,6] Supporting the use of H^+ drift to describe the transport in the hydrogen-annealing model is that similar activation energies have been measured in both hydrogen-annealing and radiation experiments.[1,4,7,8] However, there are discrepancies in other properties attributed to H^+ in the hydrogen-annealing and radiation models.[7,9] Both models assume that H^+ transport across the oxide is controlled by the bulk transport properties of H^+.

Here, we report on the effects of applied oxide field, H^+ concentration, and oxide thickness on H^+ transport in oxides of H_2 annealed Unibond material. Unlike the earlier studies of H^+ in buried oxides, our results suggest that detrapping of protons near the Si/SiO_2 interfaces has a significant effect on the transport of H^+ across the oxide. We observe that the transport time depends on the transport direction. In particular, the charge travels slower

TABLE I: Details of samples.

Sample No.	11	26	53	57	76	78	80	82
Oxide Thickness (nm)	400	400	400	400	120	170	200	400
Patterning	After Anneal	Before Anneal	Before Anneal	Before Anneal	Before Anneal	Before Anneal	Before Anneal	Before Anneal
H_2 Anneal Temperature (°C)	700	700	700	700	650	650	650	650
H_2 Anneal Time (min)	120	15	15	15	30	30	30	30
Voltage Shift (V)	46	94	56	11	4	20	34	61
Mobile Charge Density (x 10^{12} cm^{-2})	2.3	4.7	2.8	0.54	0.66	2.4	3.4	3.0

when moving from the substrate to the top silicon interface than in the opposite direction. We also find that the H^+ induced self-field in the oxide has a negligible effect on the transport time even when it is larger than the applied field, which also suggests that H^+ detrapping must contribute to its transport. Finally, in studies of the dependence of transport on oxide thickness, we find that the results cannot be described by a power law as others have previously suggested. Because of this, estimations of write times for NVFETs with thinner oxides, which have been determined assuming a power law dependence, may not be accurate.

2. EXPERIMENTAL INFORMATION

Samples are from Unibond wafers with 120, 170, 200, or 400 nm thick buried oxides. Each sample had the top Si layer patterned in 0.75 mm diameter Si dots either before or after annealing in pure H_2 at 650°C or 700°C. Anneals were conducted from 15 min to 2 h. Samples were rapidly cooled to 250°C in the H_2 ambient. Previous work has noted that a rapid cool-down in the hydrogen containing ambient is necessary to keep a high concentration of H^+ in the buried oxide layer.[7,10] The effect of the H_2 annealing was verified by examining unhydrogenated control samples.

Point contact I-V measurements were used to monitor the flow of H^+ across the oxide. Two probes placed on the top Si layer formed the source and drain while the Si substrate served as the gate. The charge sensed at the Si layer/oxide interface was measured from the midgap voltage of the I-V curve. Previous work has shown that the total charge in the oxide due to H^+ is constant for Unibond material.[7] Hence, from the midgap voltage, the centroid position can be determined. The H^+ was cycled back and forth across the oxide by alternating the polarity of the applied field. Each polarity was typically applied for 10,000 sec or more to drive nearly all of the H^+ across the oxide. Its centroid was monitored by a fast I-V technique that takes 0.4 sec and minimizes H^+ movement during the I-V sweep.[7] As the H^+ moved, there was no significant stretchout of the I-V curves. The midgap voltage was approximated by the voltage at which the I-V curve passed through 1 nA.

The details of samples described in the paper, which are typical of the many sample that we have examined, are listed in Table I. Included are the state of the top Si layer

during hydrogen annealing, the total voltage shift, the mobile charge concentration, and the annealing time. In all samples, the fixed charge in the samples is at most 4% of the mobile charge. The samples annealed with the top Si layer intact, such as number 11, in general had lower concentrations of mobile charge than those that had their top Si layer patterned before they were annealed. The conditions that affect H^+ formation in the buried oxide are complex and still under dispute.[11,12] We use the factor of 10 variation in the H^+ concentration for the samples in Table I to provide insight into the mechanisms of H^+ transport.

3. FIELD DEPENDENCE STUDIES

The first piece of evidence that leads us to conclude the interface plays a dominant role is the transport dependence on the polarity of the applied oxide field, as indicated in
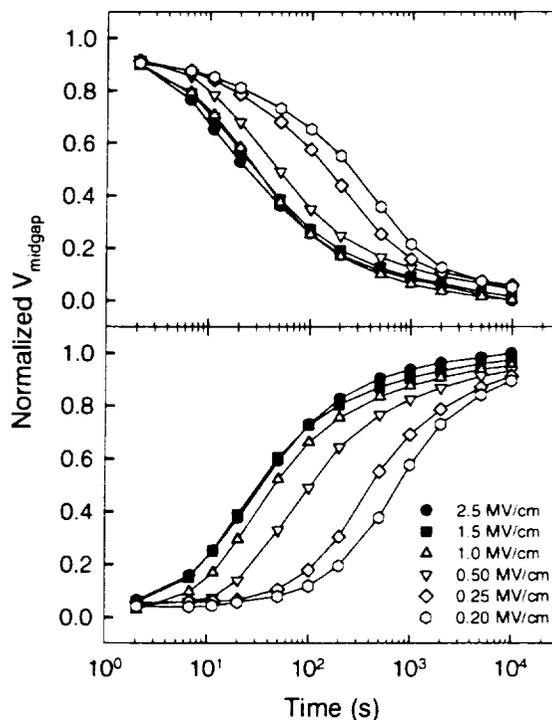

FIG. 1. Normalized midgap voltage (V_{midgap}) as a function of time (s) for sample no. 11 when the applied oxide field in negative [Fig. 1 (a)] and positive [Fig. 1(b)]. When V_{midgap} is equal to 0 and 1, the charge centroid is aligned at the substrate Si and top Si interfaces, respectively.
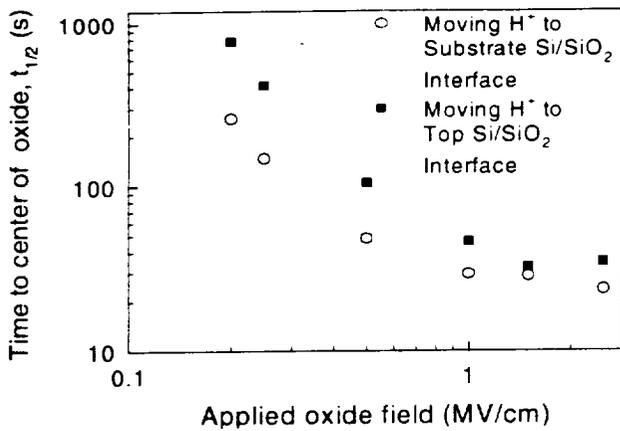
FIG. 2. Time for H$^+$ centroid of Fig. 1 to reach the center of the oxide, t$_{1/2}$ (s), as a function of the applied oxide field (MV/cm).

Figs. 1 and 2. In Fig. 1, the normalized midgap voltage is plotted as a function of time for sample 11. In Fig. 1(a) and 1(b), biases applied to the substrate are negative and positive, corresponding to H$^+$ transport from the top Si to substrate Si and substrate Si to top Si interfaces, respectively. The magnitude of the applied oxide field ranged from 2.5 MV/cm to 0.2 MV/cm. The normalized V$_{midgap}$ voltages are 0, when the H$^+$ are aligned at the substrate Si/SiO$_2$ interface, and 1, when the H$^+$ are aligned at the top Si/SiO$_2$ interface. The actual V$_{midgap}$ voltages when the charges are aligned at these interfaces are -2.0 and -46 V, respectively. The results indicated in Fig. 1 demonstrate that transport of charge is strongly affected when the applied field is less than approximately 1 MV/cm. When the applied field is larger, we begin to observe saturation in the transport dependence. By comparing Fig. 1(a) with 1(b), we noted that the charge moves slower when transported from the substrate Si/SiO$_2$ interface to the top Si/SiO$_2$ interface [Fig. 1(b)] than when it is transported in the opposing direction [Fig. 1(a)]. This effect is even more prominent at the lower magnitudes of the applied field.

The data of Fig. 2, in which the time for the centroid to reach the center of the oxide, t$_{1/2}$, is plotted as a function of the applied field, also demonstrate the difference in transport time based on the direction of charge motion. The filled square and unfilled circle data are measured when the applied fields are negative (H$^+$ moves toward the substrate Si) and positive (H$^+$ moves towards the top Si), respectively. Specifically, t$_{1/2}$ is on average 1.7 larger for moving charge from the substrate Si/SiO$_2$ interface to the top Si/SiO$_2$ interface than in the opposite direction when 0.5 MV/cm is used to cycle the charge. The magnitude of the transport asymmetry varies by 15% across the sample set shown in Table I indicating the consistency of the transport asymmetry.

Several mechanisms for moving the H$^+$ centroid across the oxide could be used to describe the effect shown in Fig. 1. We suggest a detrapping based model, which can account for the asymmetric transport times shown in Figs. 1 and 2.

In the H$^+$ drift model, interactions of H$^+$ with the local oxide environment should not be affected by the direction of the centroid movement. The longer transport times that are required to move the H$^+$ from the substrate Si/SiO$_2$ interface to the top Si/SiO$_2$ interface than in the opposite direction can be interpreted to suggest that more defects are located near the substrate Si/SiO$_2$ interface. This corresponds with the position of the bond used to create the buried oxide layer of Unibond wafers.

The saturated data of Fig. 2 do not demonstrate a power law dependence on applied field, although others have previously observed such a relationship.[1] While materials differences may contribute to the inconsistency between the results presented in this paper and that reported earlier, the discrepancy in results is likely due to differences in measurement techniques. Whereas the former study extracted transport times from the change in drain current as the gate voltage was held constant,[1] the fast I-V measurement technique used here allows for a more accurate evaluation of the position of the charge centroid in the BOX. The saturation in t$_{1/2}$ above 1 MV/cm may be due saturation in the rate of H$^+$ detrapping at this field or could indicate that at higher fields bulk transport dominates the transport characteristics.

4. H$^+$ CONCENTRATION DEPENDENCE STUDIES

The effect of H$^+$ detrapping on transport is also shown in Fig. 3, where we compared the transport of samples 26, 53, and 57, which contain difference H$^+$ densities. Figures 3(a) and 3(c) track the H$^+$ centroid as it shifts from the top Si/SiO$_2$ interface to the substrate Si/SiO$_2$ interface and Fig. 3(b) and 3(d) measure its motion in the opposite direction. If the H$^+$ in samples 26, 53, and 57 were distributed in a sheet, the field discontinuity traversing across the charge sheet would be 2.2 MV/cm, 1.3 MV/cm, and .25 MV/cm, respectively. In Fig. 3(a), 3(b), 3(c), and 3(d), the charge is cycled by applying an oxide field of -0.5 MV/cm, +0.5 MV/cm, -1.5 MV/cm, and 1.5 MV/cm, respectively. Comparing the sheet charge field with the applied fields, we see that the self-fields from the H$^+$ range from smaller to larger values than the applied field. The variations in the transport characteristics among the samples indicated in this figure are small and are less than would be expected if the self-fields influenced the H$^+$ transport.

This result indicates that H$^+$ trapping near the Si/SiO$_2$ interfaces is more important than its bulk behavior. If the motion of the centroid across the oxide were mainly dependent upon protonic drift, the H$^+$ ions would be spread across the oxide and the individual H$^+$ ions would experience a field due to the remaining H$^+$. Because this self-field is proportional to the charge density, the H$^+$ concentration would affect the transport time. However, because the transport of the H$^+$ centroid is weakly dependent on H$^+$ concentration, this figure suggests the rate-limiting step in moving H$^+$ across the oxide is
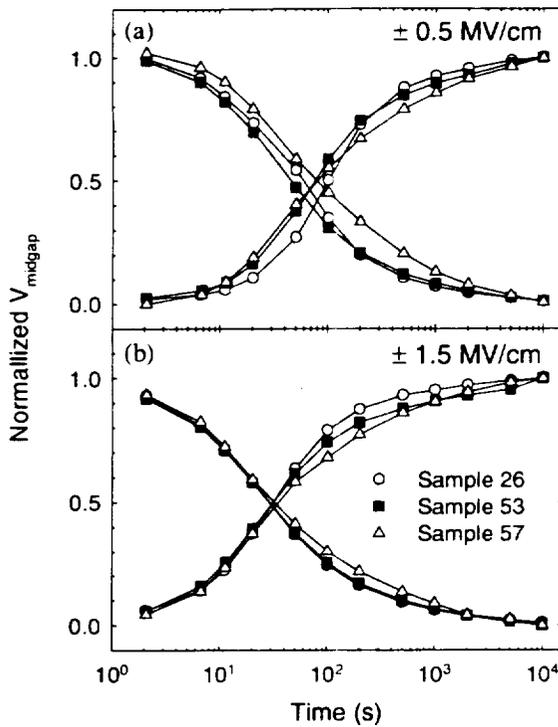
FIG. 3. Normalized midgap voltage (V_{midgap}) as a function of time (s) for sample no. 26, 53, and 57. When V_{midgap} is equal to 0 and 1, the protons are aligned at the substrate Si and top Si interfaces, respectively.

detrapping from defects near the Si/SiO_2 interfaces. In light of the lack of concentration dependence, we can neglect its effects on transport because when the H^+ are near either Si/SiO_2 interface, there is no net field applied to individual positive charges from the rest of the H^+. The model of transport controlled by detrapping from near interface defects also agrees with recent theoretical models of the buried oxides by Pantelides et al.[13] Like this paper, their work suggests the presence of defects near the Si/SiO_2 interfaces that trap the H^+ in deep wells with asymmetric barriers such that the positive charge is confined to the oxide.

5. OXIDE THICKNESS DEPENDENCE STUDIES

In Fig. 4, transit time to the center of the oxide, $t_{1/2}$, is plotted as a function of the thickness of the oxide layers, t_{ox}, for Unibond samples 76, 78, 80, and 82, which were simultaneously hydrogen annealed. The mobile charge was introduced into the buried oxide layers of these samples by annealing in H_2 at 650°C for 30 min after each of the samples had its top Si layer patterned into 0.8 mm diameter dots. The applied field used to cycle the H^+ in each of the Unibond samples was 1 MV/cm, though similar results were observed when the charge was cycled at other magnitudes of the applied field. The unfilled square and filled circle data are measured when the applied fields are negative (H^+ moves towards the top Si) and positive (H^+ moves toward the substrate Si), respectively. As is
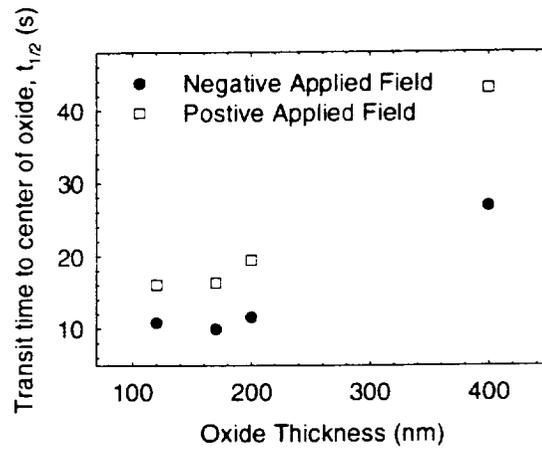
FIG. 4. Time for H^+ centroids of Unibond samples (no. 76, 78, 80, and 82) to reach the center of the oxide, $t_{1/2}$ (s), as a function of the oxide thickness (nm).

typically the case for the Unibond samples, the transport time for moving the charge from the top Si/SiO_2 interface to the substrate Si/SiO_2 interface is faster that that observed when cycling the charge in the opposite direction.

In their studies of H^+ generated by pulsed irradiation, Brown and Saks observed that the interface state formation time increased as $t_{ox}^{2.6}$.[5] Because the time during which the H^+ interacts with the hydrogen passivated Si dangling bond creating the interface state is negligible, this result indicates that H^+ drift in the oxide increases as $t_{ox}^{2.6}$. This result was found to be in agreement with models of H^+ diffusion in SiO_2, which suggests that the time for the H^+ to drift across the oxide has a power law dependence on the thickness of the oxide. The data of Fig. 4 do not demonstrate a power-law dependence on oxide thickness. Thus, this supports our previous conclusion that motion of the H^+ generated by H_2 annealing is dominated by some mechanism other than H^+ drift. Our observation is different, however, from other studies of H^+ motion in buried oxides. Vanheusden and coworkers reported that the rate of the motion of protons introduced into buried oxides by H_2 annealing increased approximately as the cube of the oxide thickness.[1] In their study, the charge was cycled in oxides of various thicknesses using a uniform bias. However, this work does not account for the differences in applied field in the different thickness oxides under −10 V bias whereas the data in Fig. 4 were all measured using ± 1.0 MV/cm to cycle the charge.

The results indicated in Fig. 4 also have implications with regard to the detrapping model of H^+ transport. One might expect if H^+ detrapping dominates H^+ transport then the time to move the charge across the buried oxide should be independent of the oxide thickness. However, as shown in Fig. 4, the H^+ traverse the thinner oxides faster. This may suggest that motion of the H^+ depends on a combination of H^+ detrapping and oxide drift. It could also suggest that the concentrations and distributions of traps in the thinner

oxides are different from that observed in the thicker ones. Because H^+ motion does not depend on H^+ drift alone, estimates for write times for thinner oxides of NVFETs can not be ascertained by extrapolation of measurements made on thicker oxides. As a result, write times reported by Vanheusden and coworkers may be over or under estimated.[10] However, no predictions of transport times from our results can be inferred for the thinner oxides of NVFETs as it is unclear what concentrations and distributions of H^+ traps would or could exist in such oxides.

6. CONCLUSION

To summarize our observations, we find that the transport of H^+ ion across a BOX depends on applied field and the direction of charge transport through the oxide, although the self-field of the charge has a negligible effect on transport even when larger than applied fields. These results suggest detrapping from near interface defects is the rate-limiting step in transporting the H^+ across the oxide. Transport occurs on a slower time scale when moving charge from the substrate Si to the top Si in the Unibond material, indicating that there are larger concentrations of defects acting as H^+ traps near the substrate Si/SiO_2 interface. Transport controlled by H^+ detrapping brings into question previous estimates of NVFET write times which were computed assuming H^+ oxide drift was the rate limiting step in moving the charge across the oxide. Because it is unknown what concentrations or distributions of traps thinner NVFET oxides will contain, predictions of write times for these devices can not be determined.

REFERENCES

[1] K. Vanheusden et al., Nature **386**, 587 (1997).

[2] W. L. Warren et al., IEEE Trans. Nuc. Sci. **44**, 1789 (1997).

[3] H. A. Kurtz and S. P. Karna, IEEE Trans. Nucl. Sci. **46**, 1574 (1999).

[4] F. B. McLean, IEEE Trans. Nucl. Sci **27**, 1651 (1980).

[5] D. B. Brown and N. S. Saks, J. Appl. Phys. **70**, 3734 (1991).

[6] N. S. Saks and D. B. Brown, IEEE Trans. Nucl. Sci. **36**, 1848 (1989).

[7] R. E. Stahlbush, R. K. Lawrence, and H. L. Hughes, IEEE Trans. Nucl. Sci. **45**, 2398 (1998).

[8] P. S. Winokur et al., J. of Appl. Phys. **50**, 3492-3494 (1979).

[9] K. Vanheusden et al., IEEE Tran. Nucl. Sci. **45**, 2391 (1998).

[10] K. Vanheusden et al., J. Non-Cryst. Solids **254**, 1 (1999).

[11] A. G. Revesz, R. E. Stalhbush, and H. L. Hughes, in *The Physics and Chemistry of SiO_2 and the SiO_2-Si Interface-4*, edited by I. J. R. B. H.Z. Massoud, M. Hirose, and E.H. Poindexter (The Electrochemical Society, Pennington, NJ, 2000), p. 235.

[12] K. Vanheusden et al., Microelectronic Engineering **48**, 363 (1999).

[13] S. T. Pantelides et al., IEEE Trans. Nuc. Sci., in press.

Functional Nanostructured Particle Arrays for Nanocrystal Nonvolatile Memories

Harry Atwater
Thomas J. Watson Laboratory of Applied Physics
California Institute of Technology

Collaborators:

Michele Ostraat
Mark Brongersma
Elizabeth Boer
Julie Casperson
Stefan Maier
Richard C. Flagan

Outline:

Si Nanoparticle MOS Structures:
- Nonvolatile Memory Devices
- Probing Defect States

Jan de Blauwe,
Martin Green

Lucent Technologies

Support: NSF, NASA, Lucent

JPL L. Douglas Bell

Caltech Nanoparticle Memory Project

Tunnel Oxide | Gate | Si Nanocrystals
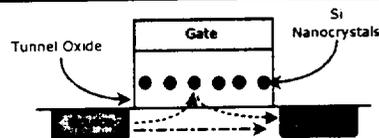
1 transistor/cell nonvolatile memory with Si nanoparticle floating gate:
•thin tunnel oxide
•fast
•greater reliability

Previous work:
(Tiwari et al *Appl Phys Lett* 68 (10) 4 March 1996)

Materials:
AFM Charging of Si Nanoparticles and Nanoscale Charge Imaging via Electrostatic Force Microscopy

Devices:
Improved Performance Nonvolatile Memory

80
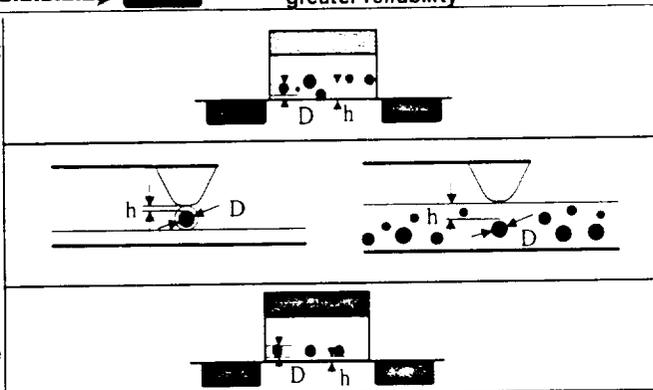
Why Small is Good: @ 300K, kT = 26 meV

GaAs/AlGaAs Dual gate device
(e-beam lithography)
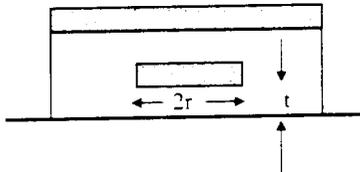
$t = 5\text{-}25$ nm
$r = 25\text{-}50$ nm
$C = 1\text{-}10$ fF
$\Delta E_{1e\text{-}} = e^2/2C = 0.05\text{-}0.5$ meV

Size-classified Si nanoparticle on SiO_2
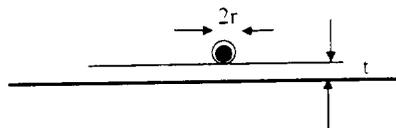
$t = 2\text{-}5$ nm
$r = 2\text{-}5$ nm
$C = 1\text{-}10$ aF
$\Delta E_{1e\text{-}} = e^2/2C = 50\text{-}500$ meV

Si nanocrystal synthesis and classification by size

silicon source rod
graphite crucible
and silicon melt
nanocrystal aerosol
excess aerosol
Turbo pump
water cooled Cu electrodes
Ar inlet
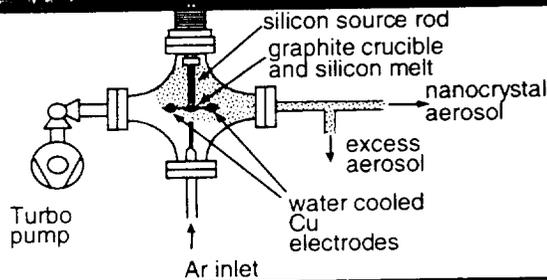
•Silicon evaporated in an atmosphere of ultra pure Ar, creating an aerosol

•Nanocrystals synthesized in a clean environment to avoid oxidation

•Size classification done with a radial differential mobility analyzer (RDMA)

•Before entering RDMA particles are charged

•After classification, particles are deposited on a Si substrate or SiO_2 film

particle trajectories   excess flow, Q_e
electric field lines   aerosol flow, Q_a
sheath flow, Q_{sh} V_{DMA} sample flow, Q_s
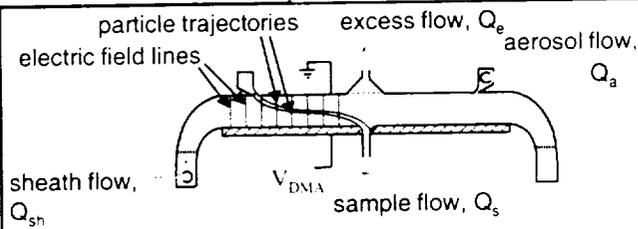
S.-H. Zhang et al. *Aerosol Science and Technology.* 23:357-372(1995)

Details of size classification

Aerosol particles (charge q)

Sheath flow

v_{e1} v_{e2} E $v_{e2} < v_{e1}$ $\Longrightarrow$ $D_{p2} > D_{p1}$

D_p D_p

10.0 V   24.5 V   54.1 V   111 V

2.8 nm
(0.6 nm)   x 7   x 4   x 3   10.8 nm
(1.8 nm)

5.2 nm   8.0 nm
(1.4 nm)   (1.6 nm)

$\Delta N/\Delta D_p$ (10^{10} nm^{-1} cm^{-2})

Nanocrystal Diameter, D_p (nm)

10 nm

10 nm

R.P. Camata et. al. *Appl. Phys. Lett.* **68** (22), 27 May 1996
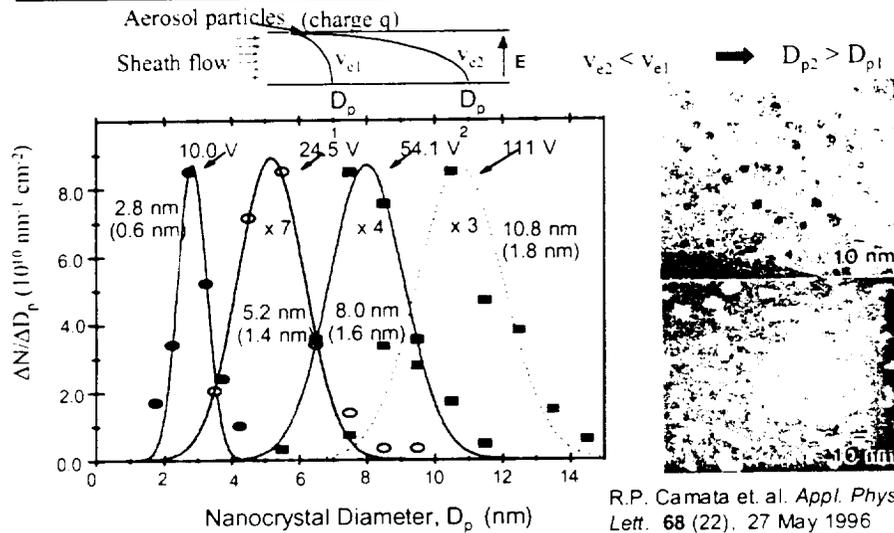
Atomic force images of silicon nanocrystals (aerosol samples)

F
(force)

contact mode
F~ 10^{-7}-10^{-6} N
r <0.5 nm

r (tip-to-sample separation)

non-contact
mode
F~10^{-12}-10^{-11} N
r~ 1-10 nm

10 nm

0

0   0.5   1 μm

Particle-substrate interactions

•weak forces: e.g. van der Waals ~10^{-11}-10^{-9} N

Where is the charge stored?

surface states?

nanocrystals?

defects in oxide?

Si/SiO₂ interface states?

E

E

E

E

E

nanoparticle/oxide interface states?
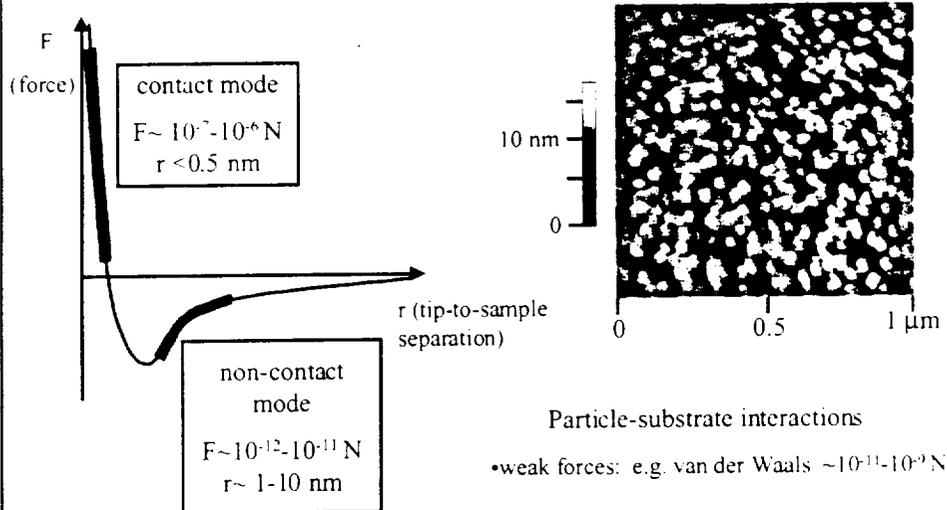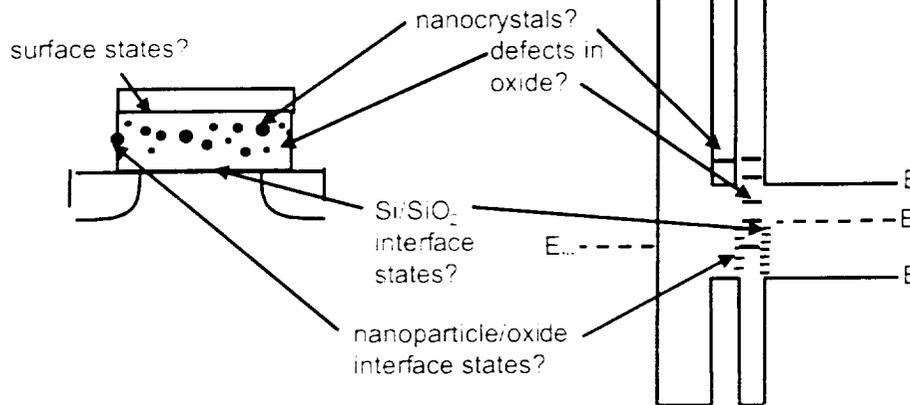
•*Charging trapping state not known in detail*
•*Transport mechanism not well characterized*

Lurking Question: Is Charge Stored in Nanoparticle Floating Gate or Oxide Defects?

t=0

t=57s

Ion Implanted Si Nanoparticle Floating Gate Material:
Stores Charge

12 nm

16 nm

0.5 μm

Similar Oxide, w/o Si Nanoparticles, but Ion Damaged by Ar⁺ Ion implantation:
No Measurable Charge Storage

1 nm

1 nm

1 μm
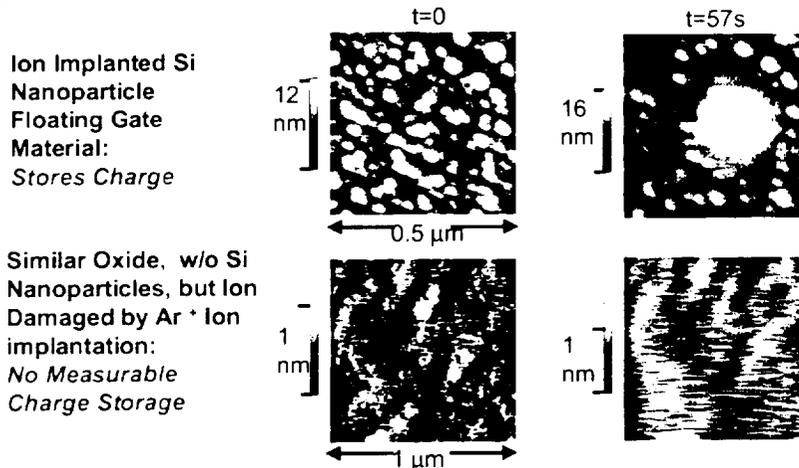
Conclusion: Locus of Charge Storage is Si <u>Nanoparticle Floating Gate</u>. Not Oxide Defects

Nanoparticle Memory Device Physics
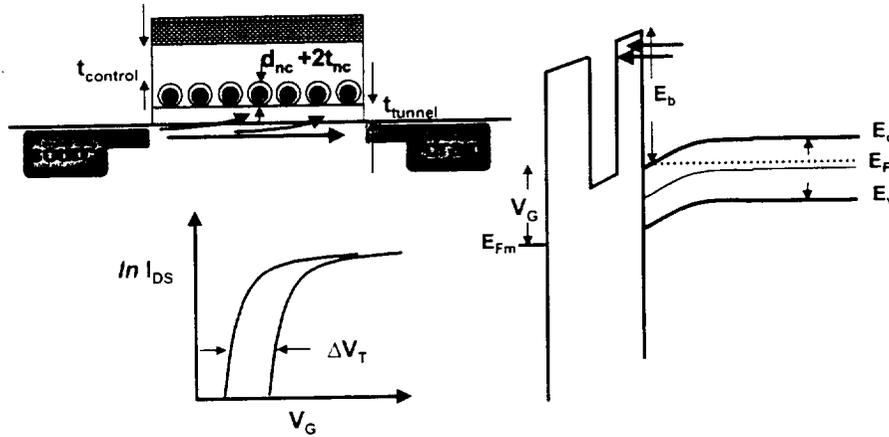
Nanoparticle Memory With Si Aerosol Floating Gate
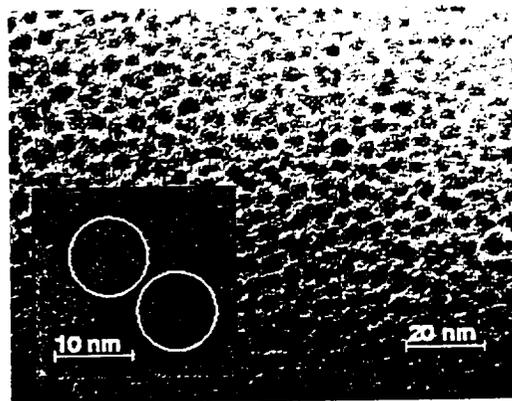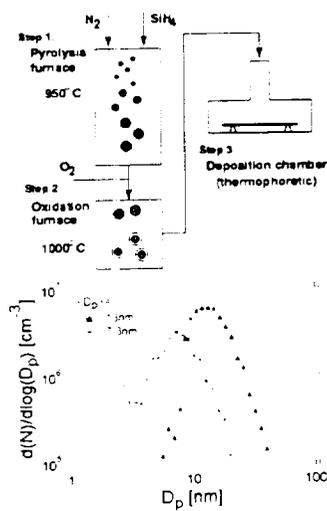
Performance of $L_{eff} = 0.2\,\mu m$ Nanoparticle Memory

Performance of $L_{eff} = 0.2\,\mu m$ Nanoparticle Memory

85
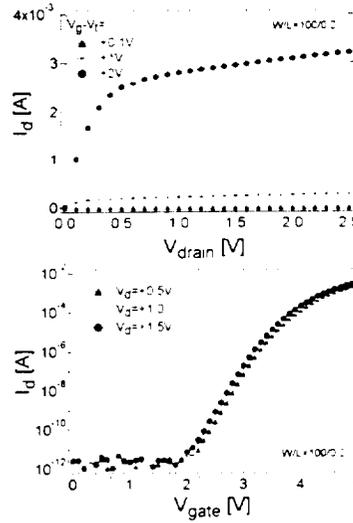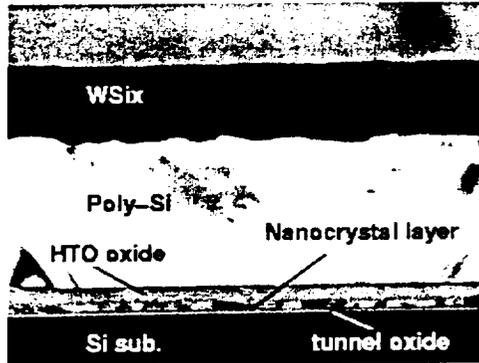
Performance of $L_{eff} = 0.2\,\mu m$ Nanoparticle Memory

Memory Cycle Durability

Memory Retention Time

P V_g=8V, V_{sub}=-5V t_p=50us
E V_g=-8V, V_{sub}=8V t_E=100ms

Program/Erase cycles [-]

all nodes grounded
V_g=2.5V V_d=1V

Retention and read disturb after cycling (10^5 cycles)
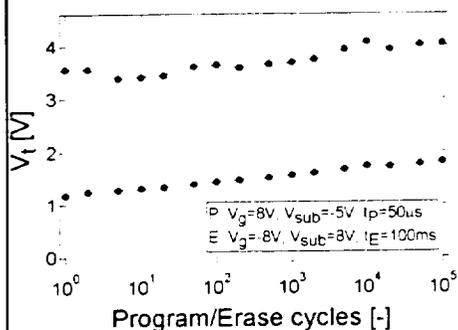
Time [s]
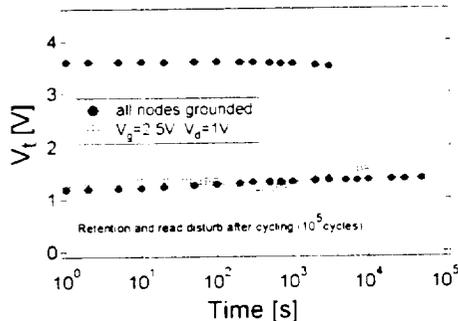
Summary:

• State-of-the-Art Nonvolatile Memory Device Performance Characteristics
• Identified Charge Storage Mechanisms in Nanocrystal Floating Gate Materials

Future:

• Spectroscopy to identify charge storage localized state energy.
• Layered Tunnel Barrier Dielectric Heterostructure for Improved Write/Erase Tunneling Performance

Quantitative Analysis of Charge Injection and Discharging of Si Nanocrystals and Arrays by Electrostatic Force Microscopy

L.D. Bell
Jet Propulsion Laboratory, Caltech

E. Boer, M. Ostraat, M.L. Brongersma, R.C. Flagan, H.A. Atwater
Caltech

Introduction / Background

- NASA requirements for computing and memory for microspacecraft emphasize high density, low power, small size, and radiation hardness.

- The distributed nature of a storage elements in nanocrystal floating-gate memories leads to intrinsic fault tolerance and radiation-hardness. Conventional floating-gate non-volatile memories are more susceptible to radiation damage.

- Nanocrystal-based memories also offer the possibility of faster, lower power operation.
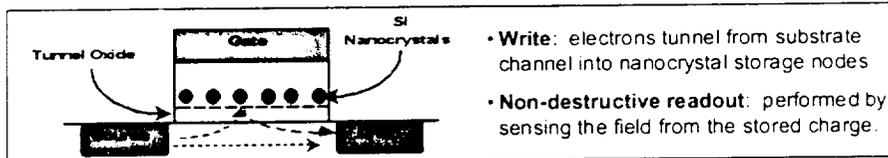
- **Write**: electrons tunnel from substrate channel into nanocrystal storage nodes

- **Non-destructive readout**: performed by sensing the field from the stored charge.

JPL

Why Small Is Good

GaAs/AlGaAs Dual gate device
(e-beam lithography)

$2r$ t
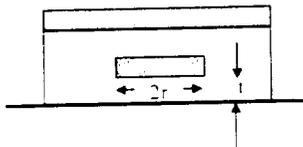
$t = 5\text{-}25$ nm
$r = 25\text{-}50$ nm
$C = 1\text{-}10$ fF
$\Delta E_{1e} = e^2/2C = 0.05\text{-}0.5$ meV

at 300K, kT = 26 meV

Size-classified Si nanoparticle on SiO

$2r$ t
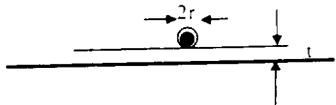
$t = 2\text{-}5$ nm
$r = 2\text{-}5$ nm
$C = 1\text{-}10$ aF
$\Delta E_{1e} = e^2/2C = 50\text{-}500$ meV

Si nanocrystal memories may enable room-temperature sensing of single-electron storage

Lucent Technologies

JPL

State of the Art

Tunnel Oxide   Gate   Si Nanocrystals
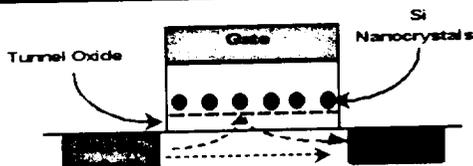
Advantages of a nanocrystal floating gate

- thin tunnel oxide - fast
- small nanocrystal size - lower power
- isolated nanocrystal floating gates - greater reliability

Previous work: (Tiwari, IBM)	Wide distribution of nanocrystal size and placement (memory element non-uniformity) D h
Single-particle addressing by AFM	h D h D
Future goal: uniform nc ensembles	Uniform nanocrystal size and placement (memory element uniformity) D h
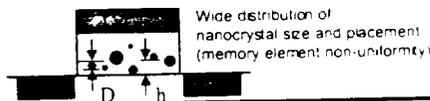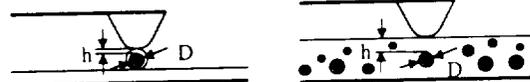
Lucent Technologies

Nanocrystal Nonvolatile Memory

JPL
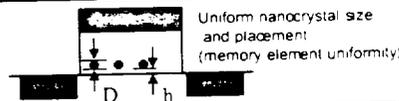
Si Nanocrystal Synthesis by Implantation

Ion implantation

35 keV Si⁺ ↓ ↓ ↓ ↓   Dose: 4×10^{16} ions/cm²

100nm SiO₂
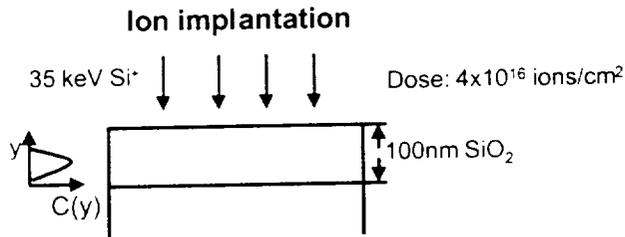
Vacuum anneal 1100°C, 10 minutes

100nm
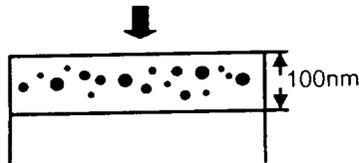
K.S. Min et al. *Appl. Phys. Lett.* **68**, 2511 (1996)

Lucent Technologies

Prepared by L. D. Bell 29 September 2000

Nanocrystal Nonvolatile Memory

JPL

Charging of Ion-implanted Si Nanocrystals

- Apparent height of charged area ranges from 11.2 nm to 4.4 nm.
- Significance: Charging discharging can be seen. Individual memory element operation can be simulated and observed.

| before charging | T=15 sec | T=58 sec |
| T=101 sec | T=187 sec | T=359 sec |

1 μm
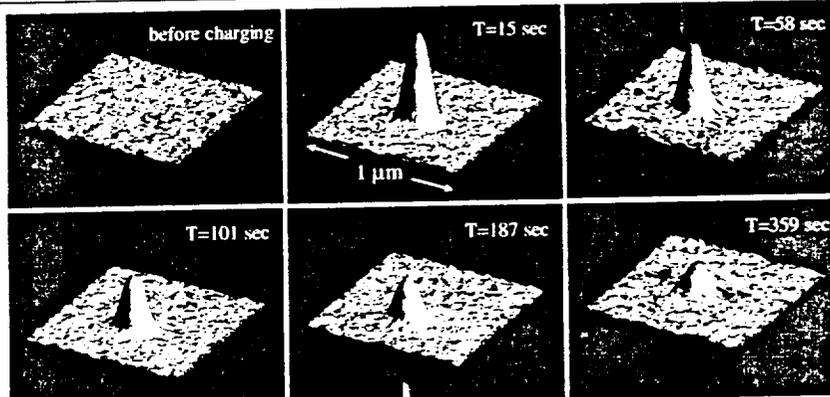
Nanocrystal charging/discharging can be imaged

Lucent Technologies

Prepared by L. D. Bell 29 September 2000

Nanocrystal Nonvolatile Memory

Charging of Ion-implanted Si Nanocrystals

JPL

Data agrees well with a model for tunneling through a field-lowered barrier
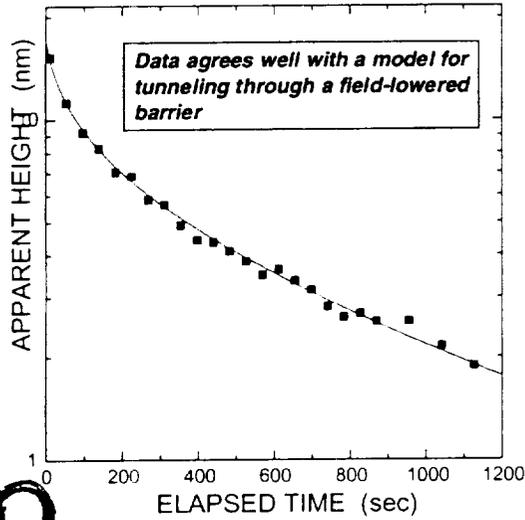
- AFM measurements can be used to track and quantify the slow discharge of a nanocrystals implanted within an oxide layer.

- Charging is measured by the AFM tip as a *localized apparent height change.*

APPARENT HEIGHT (nm)

ELAPSED TIME (sec)

Lucent Technologies

Prepared by L. D. Bell 29 September 2000

Nanocrystal Nonvolatile Memory

Charging of Ion Implanted Samples

JPL
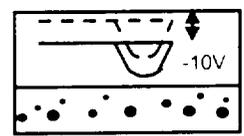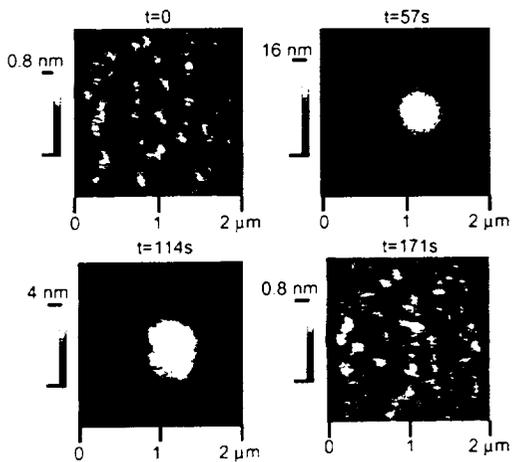
4×10^{16} at/cm^2, annealed 1100°C, 10 minutes

Where is the charge stored?
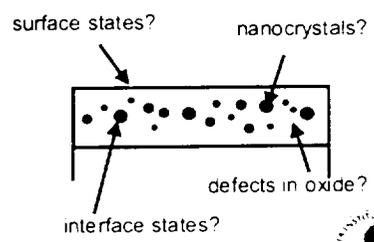
surface states?   nanocrystals?

defects in oxide?

interface states?

Lucent Technologies

Prepared by L. D. Bell 29 September 2000

Is Charge Stored in Nanoparticle Floating Gate **JPL** or Oxide Defects?

Ion-implanted
Si nanoparticle
floating gate
material:
Stores Charge
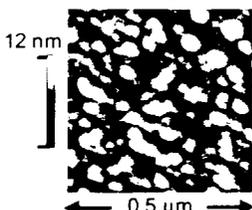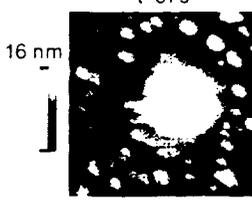
Similar oxide, no Si
nanoparticles, but ion-
damaged by Ar [+]
implantation:
*No Measurable
Charge Storage*
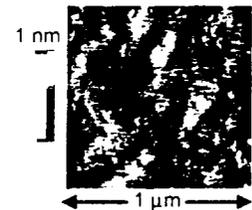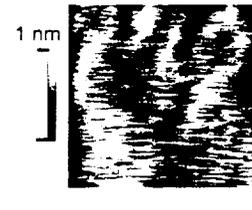
t=0   t=57s

12 nm   16 nm

← 0.5 μm →

1 nm   1 nm

← 1 μm →

*Locus of Charge Storage is
Si Nanoparticle Floating Gate, Not Oxide Defects*

Lucent Technologies

Prepared by L. D. Bell 29 September 2000

Interpretation of Charging Data **JPL**

What we want to determine:

• How much charge is stored?   *or*   *or...?*

• Distribution of charge?   *or*   *or...?*

• Dissipation of charge?   *or*   *or...?*

Model requirements:
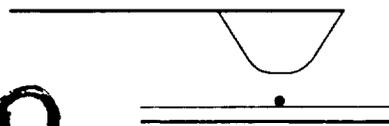
• Tip-sample convolution

• Tip-charge interaction vs.
  normal AFM operation

Lucent Technologies

Prepared by L. D. Bell 29 September 2000

JPL

Model: AFM Tip-Charge Distribution Interaction

Trapped charge q_{si} at (x_{si}, y_{si}, z_s)

q_{tj} at (x_{tj}, y_{tj}, z_{tj})

AFM Tip

D_{ij}

R_1

z

$(0,0,0)$

Method of Images

Substrate
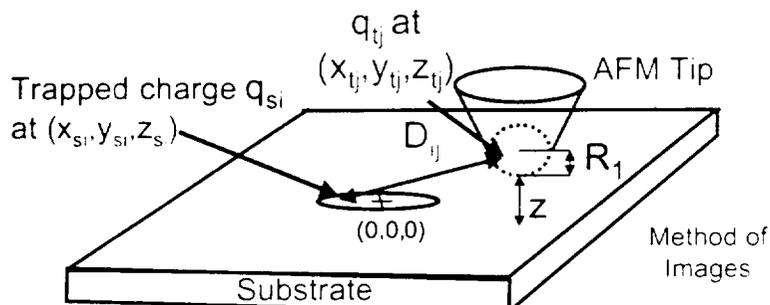
$$D_{ij}^2 = (z_{s_i} - z_{t_j})^2 + (x_{s_i} - x_{t_j})^2 + (y_{s_i} - y_{t_j})^2$$

Coulomb interaction between q_s and q_t

$$\vec{F}_{ij} = \frac{q_{s_i} {}^* q_{t_j}}{4\pi\varepsilon_o D_{ij}^2} \hat{d}_{ij}$$

Lucent Technologies

Prepared by L. D. Bell 29 September 2000

Non-Contact AFM Maps Height at Constant Force Gradient

JPL

Coulomb interaction between q_s and q_t

$$\vec{F}_{ij} = \frac{q_{s_i} {}^* q_{t_j}}{4\pi\varepsilon_o D_{ij}^2} \hat{d}_{ij}$$

Find z-component of electrostatic force; differentiate by z:

z

Finally, total force gradient is sum of electrostatic and Van der Waals interactions

$$\left(\frac{\partial F_z}{\partial z}\right)_{tot} = \left(\frac{\partial F_z}{\partial z}\right)_{vdW} + \sum_{i,j}\left(\frac{\partial F_z}{\partial z}\right)_{ij} = const$$

Lucent Technologies

Prepared by L. D. Bell 29 September 2000

JPL

Some Assumptions: Buried NC Charging

- Charge distribution: uniform, coarse grid, only on surface -- appropriate for these samples

- Neglect image charge in substrate --> Good approximation due to distances involved

- Neglect topography

- Neglect effect of hydrodynamic damping

- Assume tip is not "tapping" the surface--valid for a limited operating range

50 nm

Lucent Technologies

JPL
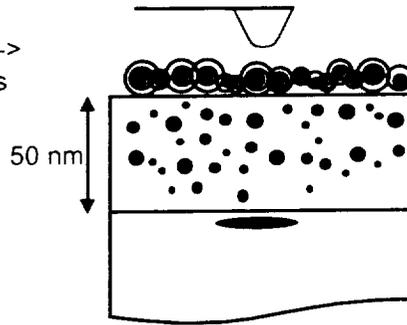
Quantitative Nanoscale Charge Imaging

Experimental EFM Image of Charged Si Nanoparticle Floating Gate:

Fit results:
- total charge = 200 electrons
- σ = 140 nm

Data
Fit

μm)

Image Simulation (w/ Tip Convolution Correction
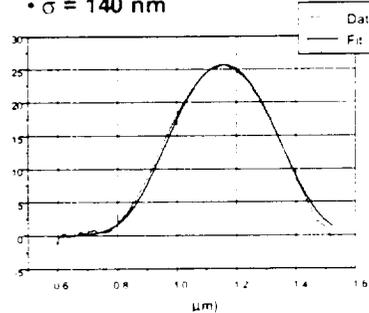
Amount of stored charge can be determined

Lucent Technologies

NASA | **Goal of electrostatic modeling** | **JPL**

- Calculated "images" qualitatively match experimental results

- Quantitative comparison will allow the amount of charge deposited, charge distribution and discharging mechanism to be determined from AFM images
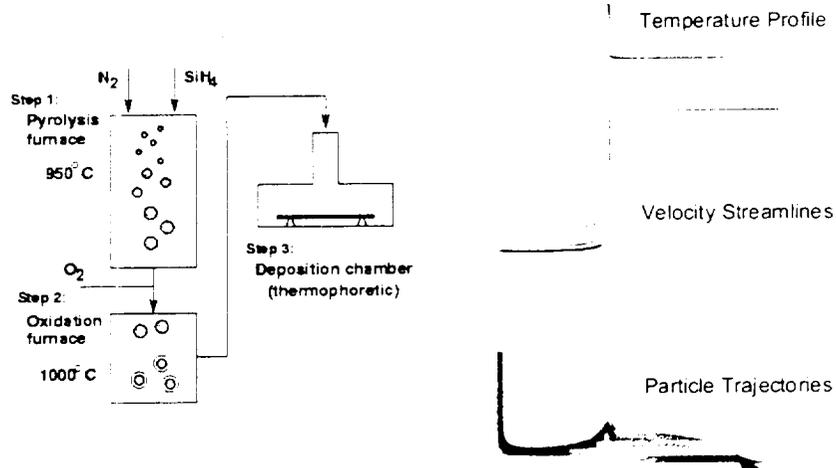
q=800e, s=80nm

16 nm

t=57s
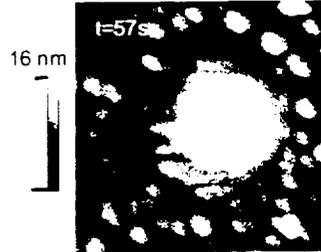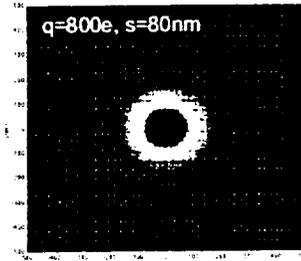
NASA | **Aerosol Synthesis of Si Nanocrystals** | **JPL**

Temperature Profile

N$_2$ SiH$_4$

Step 1:
Pyrolysis
furnace

950°C

Step 3:
Deposition chamber
(thermophoretic)

Velocity Streamlines

O$_2$

Step 2:
Oxidation
furnace

1000°C

Particle Trajectories

Aerosol synthesis and thermophoretic deposition can produce uniform layers of Si nanocrystals.

94

JPL

Nanocrystal Deposition
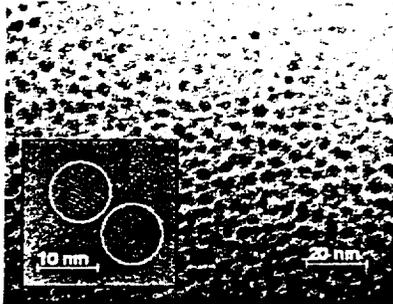
Planar view TEM of an aerosol nanocrystal monolayer.

Crystal size = 4-5nm

Particle density = 6×10^{12} cm^{-2}

10 nm

20 nm

Spherical, crystalline nanoparticle layers with tight size control and good areal coverage have been obtained
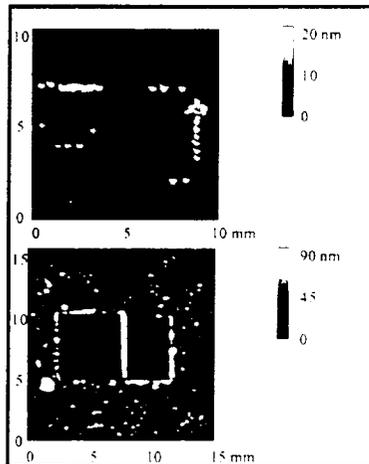
Lucent Technologies

Prepared by L. D. Bell 29 September 2000
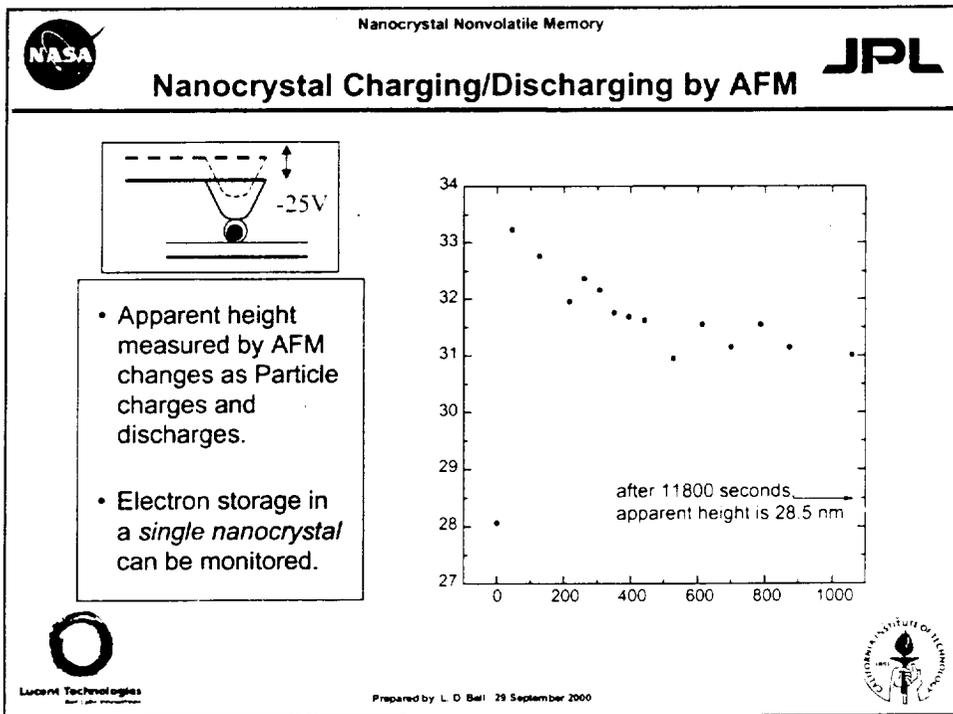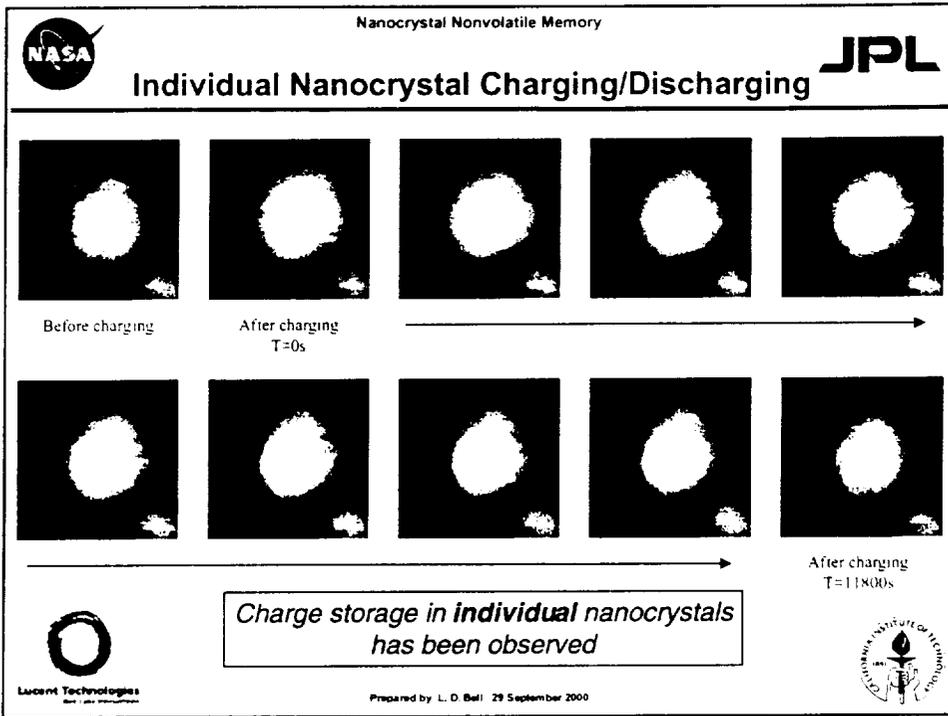

JPL

Particle manipulation with an AFM

- *Image in tapping mode*

- *Manipulate in contact mode*

20 nm

10

0

10 mm

90 nm

45

0

15 mm

Lucent Technologies

Prepared by L. D. Bell 29 September 2000

Individual Nanocrystal Charging/Discharging

JPL

Before charging

After charging
T=0s

After charging
T=11800s

Charge storage in *individual* nanocrystals
has been observed

Lucent Technologies

Prepared by L. D. Bell 29 September 2000

Nanocrystal Charging/Discharging by AFM

JPL

-25V

- Apparent height measured by AFM changes as Particle charges and discharges.

- Electron storage in a *single nanocrystal* can be monitored.

after 11800 seconds,
apparent height is 28.5 nm

Lucent Technologies

Prepared by L. D. Bell 29 September 2000

Nanocrystal Nonvolatile Memory

AFM Tip-Charge Distribution Interaction

(aerosol samples)

Now include topography
(single particle)

q_{tj} at
(x_{tj}, y_{tj}, z_{tj})

AFM Tip

Trapped charge qs
at (x_s, y_s, z_s) on
particle

D

R_1

z

(0,0,0)

Substrate

Prepared by L. D. Bell 29 September 2000

Nanocrystal Nonvolatile Memory
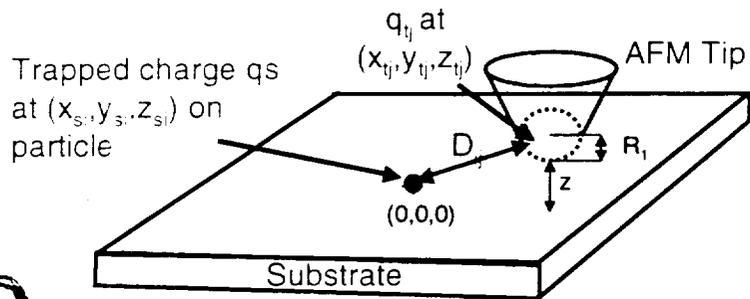
AFM Tapping-Mode Modeling

$$F = \begin{cases} F_{JKR} & \text{in contact} \\ F_{vdw} & \text{out of contact} \end{cases}$$

JKR   van der Waals

Force (nN): 800, 400, 0

tip-sample distance (nm): -1, 0, 1, 2, 3, 4

Tip position (nm): 100, 50

Time (ms): 0, 5
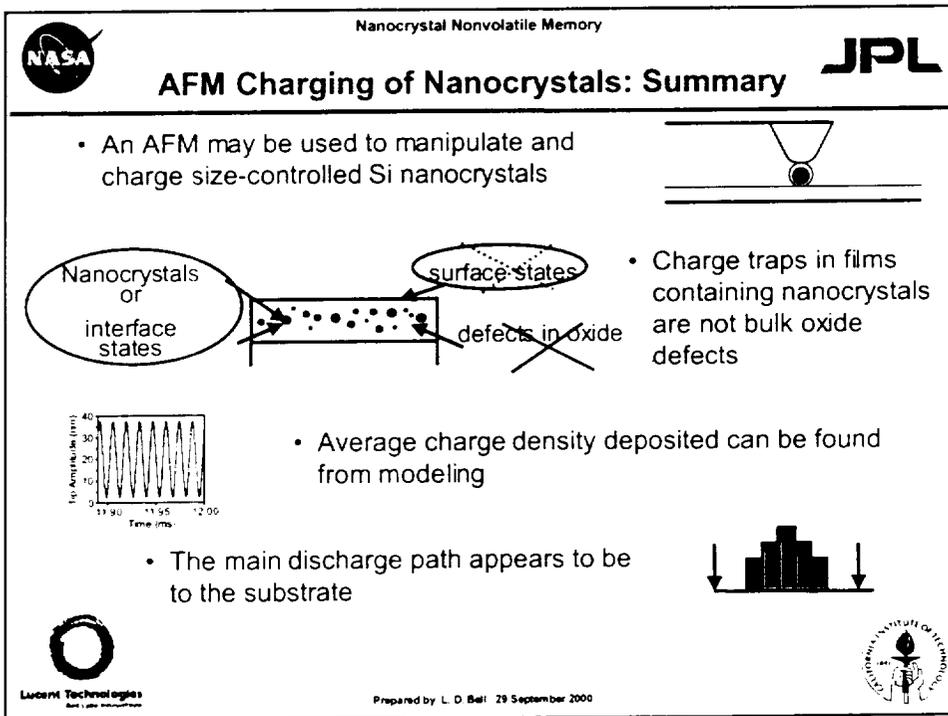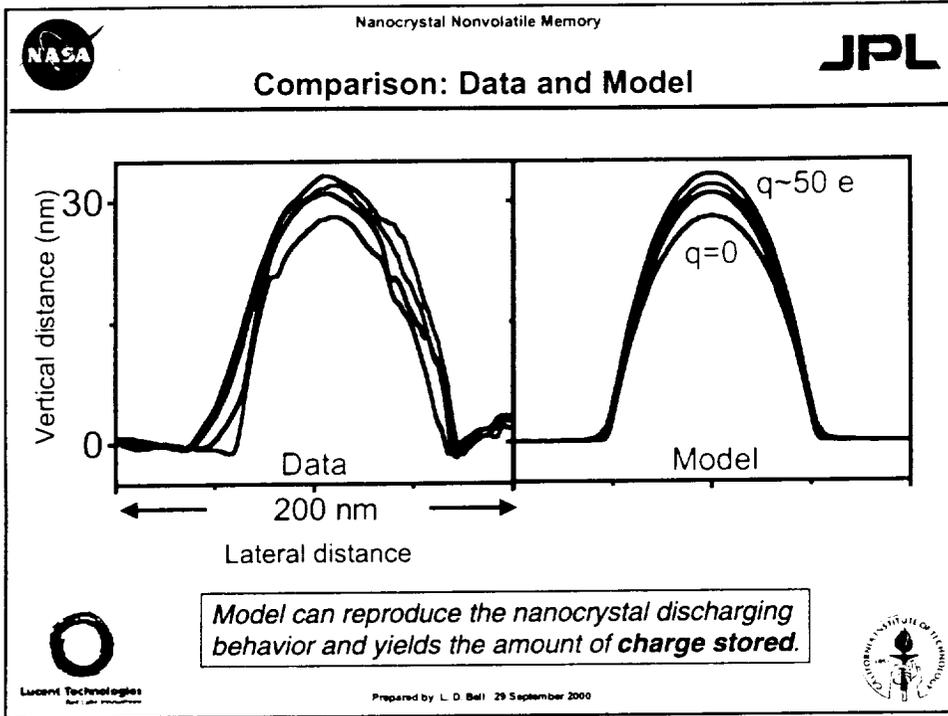
$$m\ddot{z} + \gamma\dot{z} + k(z - z_o) = F_o \cos(\omega t) + \{F_{tip-sample}\}_z$$

Prepared by L. D. Bell 29 September 2000

97

JPL

Comparison: Data and Model

q~50 e

q=0

Data

Model

200 nm

Lateral distance

*Model can reproduce the nanocrystal discharging behavior and yields the amount of **charge stored**.*

JPL

AFM Charging of Nanocrystals: Summary

- An AFM may be used to manipulate and charge size-controlled Si nanocrystals

Nanocrystals or interface states

surface states

defects in oxide

- Charge traps in films containing nanocrystals are not bulk oxide defects

- Average charge density deposited can be found from modeling

- The main discharge path appears to be to the substrate

98

JPL

Conclusions

- Si Nanocrystal charging has been accomplished with a conducting-tip AFM

- Both individual nanocrystals on an oxide surface and nanocrystals formed by implantation have been charged.

- Discharging is consistent with tunneling through a field-lowered oxide barrier

- Modeling of the response of the AFM to trapped charge has allowed estimation of the quantity of trapped charge.

- Initial attempts to fabricate competitive nanocrystal non-volatile memories have been extremely successful.

Lucent Technologies

Prepared by L. D. Bell 29 September 2000

Advanced Compact Holographic Data Storage System

Tien-Hsin Chao	(818)354-8614	Tien-Hsin.Chao@jpl.nasa.gov
Hanying Zhou	(818)354-0502	Hanying.Zhou@jpl.nasa.gov
George Reyes	(818)393-4820	George.F.Reyes@jpl.nasa.gov

Jet Propulsion Laboratory

California Institute of Technology

4800 Oak Grove Dr. Pasadena, Ca 91109

Abstract— JPL, under current sponsorship from NASA Space Science and Earth Science Programs, is developing a high-density, nonvolatile and rad-hard Advanced Holographic Memory (AHM) system to enable large-capacity, high-speed, low power consumption, and read/write of data in a space environment. The entire read/write operation will be controlled with electro-optic mechanism without any moving parts. This CHDS will consist of laser diodes, photorefractive crystal, spatial light modulator, photodetector array, and I/O electronic interface. In operation, pages of information would be recorded and retrieved with random access and high-speed. The nonvolatile, rad-hard characteristics of the holographic memory will provide a revolutionary memory technology to enhance mission capabilities for all NASA's Earth Science Mission.

In this paper, recent technology progress in developing this CHDS at JPL will be presented.

TABLE OF CONTENTS

1. INTRODUCTION

NASA's future missions would require massive high-speed onboard data storage capability to support both Earth Science and Space Science missions. With regard to Earth science observation, a 1999 joint Jet Propulsion Laboratory and Goddard Space Flight Center (GFSC) study ("The High Data Rate Instrument Study" [1]) has pointed out that the onboard science data (collected by high date rate instruments such as hyperspectral and synthetic aperture radar) stored between downlinks would be up to 40 terabits (Tb) by 2003. However, onboard storage capability in 2003 is estimated at only 4 Tb that is only 10% of the requirement. By 2006, the storage capability would fall further behind that would only be able to support 1% of the onboard storage requirements.

For Space Science, the onboard data storage requirements would be focused on maximizing the spacecraft's ability to survive fault conditions (i.e. no loss in stored science data when spacecraft enters the "safe mode") and autonomously recover from them during NASA's long-life and deep space missions. This would require the development of non-volatile memory. In order to survive in the challenging environment during space exploration missions, onboard memory requirements would also include: survive a high radiation environment (1 Mrad), operate effectively and efficiently for a very long time (10 years), and sustain at least a billion (10^9) write cycles.

Therefore, memory technologies requirements of NASA's Earth Science and Space Science missions are: large capacity, non-volatility, high-transfer rate, high radiation resistance, high storage density, and high power efficiency.

Current technology, as driven by the personal computer and commercial electronics market, is focusing on the development of various incarnations of Static Random Access Memory (SRAM), Dynamic Random Access Memory (DRAM), and Flash memories. Both DRAM and SRAM are volatile. Their densities are approaching 256 Mbits per die. Advanced 3-D multichip module (MCM) packaging technology has been used to develop solid-state recorder (SSR) with storage capacity of up to 100 Gbs [3]. The Flash memory, being non-volatile, is rapidly

Table I. A comparative chart for memory specifications

Feature	SRAM	DRAM	Flash	Holographic
Non-volatile	No	No	Yes	yes
Organization (bits/die)	512k x8 (rad hard)	16Mx8	16Mx8, 32Mx8	10 Gb/cm^3
Data Retention (@70^0 C	N/A	N/A	> 10 yrs	> 10 yrs
Endurance (Erase/write Cycles)	unlimited	unlimited	10^6 (commercial)	unlimited
Access Time, t (acc)	< 10 ns	< 25 ns	50 ns	1 ms/page
Data Transfer rate	> 800 Mb/sec	> 320 Mb/sec	160 Mb/sec	> 1 Gb/sec
Radiation (Total Dose)	1 Mrad	< 50 kRads	< 30 kRads	> 1 Mrad
Power	1 Gb/watt	1 Gb/watt	10 Gb/watt	10 Gb/watt
Package	4 Mb (die stacking) 100's Gb (Multichip Module)	128 Mb (per die) 100's Gb (MCM)	128 Mb & 256 Mb (per die) 100's Gb (MCM)	10 Gb/cm^3 cube. 1 Tb/card

gaining popularity. Densities of flash memory of 256 Mbits per die exist today. High density SSR could also be developed using the 3-D MCM technology. However, Flash memory is presently faced with two insurmountable limitations: Limited endurance (breakdown after repeated read/write cycles) and poor radiation-resistance (due to simplification in power circuitry for ultra-high density package).

It is obvious that state-of-the-art electronic memory could not satisfy all NASA mission needs. It is necessary to develop new memory technology that would simultaneously satisfy non-volatility, rad hard, long endurance as well as high transfer rate, low power, mass and volume to meet all NASA mission needs.

The comparative specifications of the holographic memory (design goal) and state-of-the-art electronic memory are listed in Table I. As shown in Table I, the holographic memory technology, upon full development, not only is simultaneously non-volatile, high-speed and rad hard, but also superior in power and volume/mass efficiency over its electronic counterpart.

2. HOLOGRAPHIC DATA STORAGE TECHNOLOGY

The Advanced Holographic Memory (AHM) system will store data in large number of holograms inside of a photorefractive crystal. Holograms will be generated by recording the light interference pattern formed by an object laser beam carrying a page of optically modulated data (image or binary bits) and a reference laser beam in a cubic photorefractive crystal. Since these images are stored in the Fourier domain and recorded in three dimensions, massive redundancy is built into the holograms such that the stored holograms would not suffer from imperfections in the media or point defects. The LiNbO$_3$ photorefractive crystal has been the most mature recording material for holographic memory due to its uniformity, high E-O coefficient, high photon

sensitivity, and commercial availability. One unique advantage for using holographic data storage is its rad hard capability. Holograms stored in photorefractive crystal have been experimentally proven to be radiation-resistant. In a recent NASA LDEF (Long-Duration Exposure on Active Optical Components) experiment, Georgia Institute of Technology (GIT), under NASA Contract NAS1-15370, has flown a Lithium Niobate holographic memory in space. The retrieved crystals only suffered surface damage and still retained their photosensitivity for hologram recordings.

With respect to the lifetime of stored data in the LiNbO$_3$ material: without a thermal fixing process after recording, data stored LiNbO$_3$ material would only have up to 1 month of shelf life. With thermal fixing, the stored data will stay unchanged for many years. Recently, a technology breakthrough that would extend storage lifetime of photorefractive memory from months to decades or longer has been achieved by researchers at Caltech and other institutions [4-5]. A new two-photon recording material, doubly doped Fe:Mn: LiNbO3 has been developed. This material possesses deep traps partially filled with electrons and shallow (intermediate) traps to trap photo-generated electrons with very long lifetime. The doubly doped extrinsic dopants (Fe^{+2}, Mn^{+2}) would provide this intermediate state. As illustrated in Figure 1a and 1b:
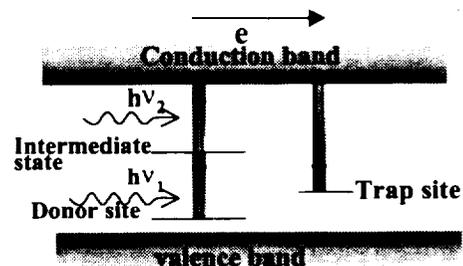

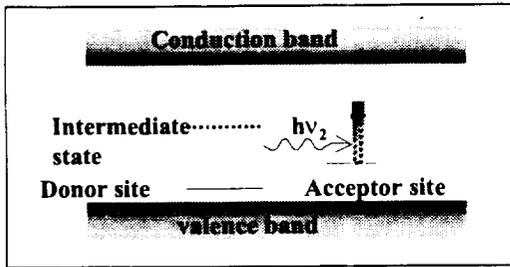Figure 1a. Nonvolatile hologram recording

Figure 1b. Nonvolatile hologram readout

During recording, a first photon (from an ultraviolet light source) is used to excite an electron from the Valence band to an intermediate state. A hologram writing photon is then used to bring the electron up to the Conduction band. The electron will then migrate and get trapped to record the interference pattern. During readout, the laser beam will readout the hologram but is with insufficient energy to elevate the electron to the conduction band. Hence the stored hologram will not be erased during the process.

Recently, we have developed a bench-top AHM Data Storage breadboard and, for the first time, demonstrated memory retrieval at video rate for both grayscale image and binary data. A 1000-page long video of images of Asteroid Toutatis were recorded and retrieved with very high fidelity. The AHM breadboard used acousto-optic (AO) scanning for multiple holograms recording and readout. This initial development has demonstrated the high-fidelity, high-speed data storage capability. A photo of the AO based AHM system and its schematic diagram is shown in Figure 2.
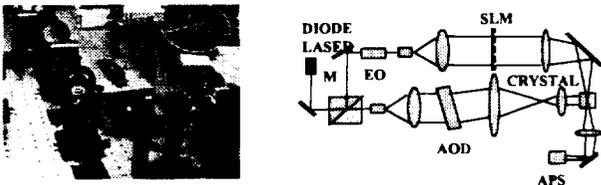
2. Photograph of the JPL developed AO- based AHM system and its corresponding schematic diagram

An experimental demonstration of the grayscale image recording and retrieving capability has recently been performed at JPL. Some Samples of high-fidelity retrieved images of the asteroid Toutatis is shown in Figure 3.
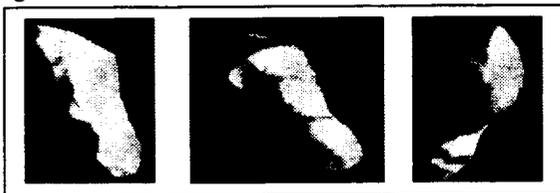

Figure 3. Grayscale images of meteoroid retrieved from JPL developed holographic memory

This JPL-developed proof-of-concept acousto-optic-scanner-based holographic memory system has demonstrated excellent performance of holographic memory. This breadboard is currently being used for bit error rate analysis of the retrieved holographic data.

In order to further miniaturize the AHM breadboard, the AO beam steering devices will be replaced by a liquid crystal beam steering spatial light modulator (BSSLM). This device will provide high-resolution beam steering capability to enable the storage/retrieval of more than 10,000 pages of holographic memory. By coupling the BSSLM with a VCSEL array, the storage density will be increased further by more than two orders of magnitude. The introduction of the BSSLM spatial multiplexing scheme would further ensure that the system is compact and consumes low power.

3. ADVANCED HOLOGRAPHIC MEMORY USING LIQUID CRYSTAL BEAM STEERING DEVICES

The proposed holographic memory architecture, as shown in Figure 4, consists of a writing module for multiple holograms recording and a readout module for readout. The writing module includes a laser diode as the coherent light source, a pair of cascaded beam steering Spatial Light Modulators (BSSLM), one transmissive and one reflective in each pair, for angular multiplexed beam steering, an input SLM for electronic-to-optical data conversion, two cubic beam splitter for beam forming, and a photorefractive crystal for hologram recording. The readout module also shares this photorefractive crystal. The readout module includes a laser diode with the same wavelength as the writing one, a pair of cascaded BSSLMs to generate phase conjugated readout beam (i.e. the readout beam is directed in opposite direction of that of the writing beam), the photorefractive crystal, a cubic beam splitter, and a photodetector array for recording the readout holograms. The system uses a angular multiplexing scheme to store multiple holograms and phase-conjugated beams to readout each holograms.

In hologram writing, the collimated laser beam (top left in Fig. 4) splits into two parts at the first cubic beam splitter: 1) The horizontally deflected light will travel across the second cubic beam splitter to read out the input data after impinging upon the Data SLM. This data-carrying beam will then be reflected into the PR crystal as the object beam; 2) The other part of the split incoming beam, travel vertically up, will first pass a transmissive BSSLM and then reflected to the reflective BSSLM. Both BSSLMs are modulated with a 1-D blazed phase gratings that are capable of beam steering with an angular deflection determined by the grating periods. By cascading two BSSLMs in orthogonal, a 2-D beam steering scheme can be achieved. This deflected laser beam will then be directed toward the PR crystal as the reference-writing

beam. It will meet the object beam inside the PR crystal to form an interfering grating (hologram). Each individual hologram is written with a unique reference angle and can only be readout at this angle (or its conjugated one). By varying the reference beam angle in sequential recording, very large number of holograms can be recorded in the recording medium.
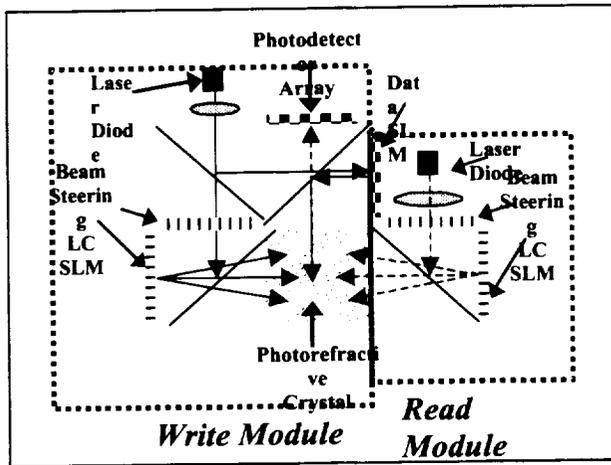

Figure 4. System schematic architecture of an Advanced Holographic Memory

For hologram readout, we have devised an innovative phase conjugation architecture. This phase conjugation scheme will enable lensless hologram readout with minimal distortion (low bit error rate). As shown in Fig.4, a second pair of transmissive and reflective BSSLMs combination will be used to provide a phase-conjugated readout beam (with respect to the writing reference beam). After the beam impinges upon the PR crystal, the diffracted beam from the recorded hologram will exit the PR crystal backtracking the input data beam path, due to phase-conjugation property. It will then directly impinge upon the photodetector array. Thus the reconstructed data will be recorded with excellent quality.

4. BEAM STEERING SPATIAL LIGHT MODULATOR

JPL has recently collaborated with the Boulder Nonlinear System Co. (BNS) to develop a BSSLMS. This device is built upon a VLSI back plane in ceramic PGA carrier. A 1-dimensional array of 4096 pixels, filled with Nematic Twist Liquid Crystal (NTLC), is developed on the SLM surface. The device aperture is of the size of 7.4 mm x 7.4 mm, each pixel is of 1 mm x 7.4 mm in dimension. Currently, the response time can reach 200 frames/sec. In future, by replacing the NTLC with Ferroelectric Liquid Crystal (FLC), the speed may be increased by one order of magnitude (i.e. > 2000 frames/sec). A photo of this BSSLM is shown in Figure 5.
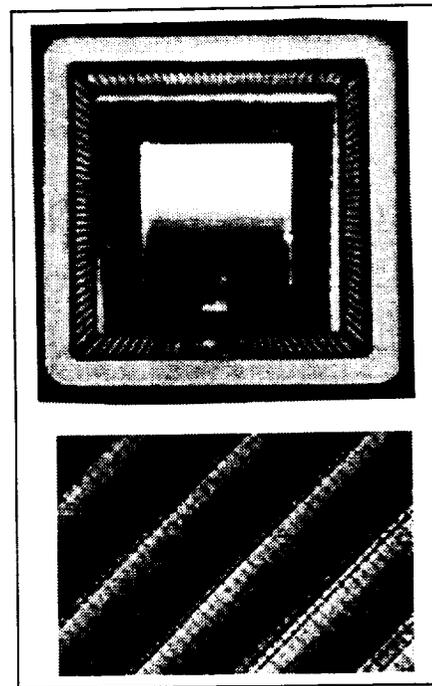

Figure 5 a photograph of a 1 x 4094 Beam Steering Spatial Light Modulator and a magnified view of the grating structure over SLM surface.

The principle of operation of this BSSLM is illustrated in Figure 6. This SLM is a phase-modulation device. In operation, a driving signal in the form of a linear ramp function will be applied to the SLM. The corresponding phase profile over the SLM will be a quantized multiple-level phase grating with its phase profile modulated from 0-to-2π periodically. A readout laser beam, upon reflected from this SLM, will be deflected by an angle q. This angle can be described as:

$$\theta = \sin^{-1}(\lambda/d)$$

Where λ is the wavelength of the laser beam. Thus, beam steering can be achieved by varying the period of the phase grating
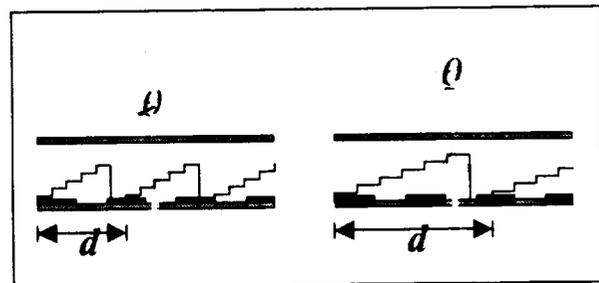

Figure 6. Beam steering using a phase modulation SLM with variable grating period.

The diffraction efficiency, η, of this device is

103

$$\eta = \left(\frac{\sin\left(\pi/n\right)}{\pi/n} \right)^2$$

Where n: number of steps in the phase profile. For example $\eta \sim 81\%$ for $n = 4$, and $\eta \sim 95\%$ for $n = 8$.

Number of resolvable angles can be defined by:

$$M = 2m/n + 1$$

Where m is the pixel number in a subarray, and n is the minimum number of phase steps used. For example, $M = 129$ for $m = 512$, $n = 8$ with a 1 x 4096 beam steering device.

Primary advantages of using such a electro-optic beam steering device for angular multiplexing during holographic data recording will include: No mechanical moving parts; Randomly accessible beam steering; Low voltage / power consumption; Large aperture operation;

The massive storage capability of holographic memory has been demonstrated by Caltech researchers. It was reported that up to 160,000 pages (i.e. 160 Gb of memory) of hologram were stored in a LiNbO$_3$ PR crystal with 1 cm^3 volume using a scanning mirror to create angular multiplexing for each reference beam. However, the scanning mirror scheme that would require mechanically controlled moving parts is not suitable for space flight. Thus we are currently developing an all electro-optic controlled angular multiplexing scheme with high-speed, high-resolution without having to use any moving parts. We are currently utilizing an all-phase beam steering device, the BSSLM.

The current transimissive BSSLM device is a 1 x 1024 array with resolvable spots about 64. The reflective BSSLM device is a silicon-based 1-D diffractive beam steering device. The current device is a 1 x 4096 array, which has about 128 resolvable spots. The total resolvable spots from these cascaded BSSLMs would be around 11,520. Thus by using two pairs of such cascaded BSSLMs for beam steering, a total of more than 10,000 pages of hologram can be stored and retrieved in a single

cubic centimeter of PR crystal. Since each page can store about 1000 x 1000 pixels of data (1 Mbytes), the total storage capacity will reach 10 Gigabytes.

5. CONCLUSIONS

In summary, the design goals of the proposed holographic memory are: high storage capacity (10 Gigabytes per module), high-speed (1ms per page), and low power (100mW laser power for writing, 10mW for readout).

Unique advantages of this holographic memory system over that of the state-of-the-art counterparts are: a) The need of only a laser diode as the writing source (the state-of-the-art laser diode can provide power of 100mW) will lend to a miniature packaging of the whole system; b) The use of a pair of BSSLMs for angular multiplexing will provide for a storage capacity of 10,000 pages of hologram (10GB) in a 1 cubic centimeter PR crystal without using moving parts; c) The use of an innovative phase-conjugation readout scheme will make possible high-fidelity data retrieval.

6. ACKNOWLEDGEMENT

7. REFERENCES

1. K. F. Strauss, T. Daud, "Overview of radiation tolerant unlimited write cycle nonvolatile memory", Proceeding of IEEE Aerospace Conference 2000.
2. W. Schober, F. Lansing, K. Wilson, and E. Webb, "High Data Rate Instrument Study", JPL publication 99-4. Semiconductor memories, A. K. Sharma, IEEE Press, 1996.
3. L. Hesselink et al, Science Vol.282, p1089, 1998.
4. D. Psaltis et al, Opt. Lett. 24, p652, 1999
5. Tien-Hsin Chao, Hanying Zhou, and George Reyes, "Highdensity holographic memory and its applications to optical pattern recognition," SPIE

Proceeding of International Symposium on Optical
Data Storage, Kauai, Hawaii, (July, 1999)

6. H. Zhou, F. Zhao, and F. T. S. Yu, "Effects of
   recording-erasure dynamics on the storage capacity
   of a wavelength-multiplexed reflection-type
   photorefractive hologram," *Appl. Opt.* 33, 4339-44
   (1994).

Ultra-high density holographic memory module
with solid-state architecture

Vladimir B. Markov,

MetroLaser Inc., 18010 Skypark Circle, Suite 100, Irvine, CA 92614

Tel.: (949)-553-0688. E-mail: vmarkov@metrolaserinc.com

Abstract - NASA's terrestrial, space, and deep-space missions require technology that allows storing, retrieving, and processing a large volume of information. Holographic memory offers high-density data storage with parallel access and high throughput. Several methods exist for data multiplexing based on the fundamental principles of volume hologram selectivity. We recently demonstrated that a spatial (amplitude-phase) encoding of the reference wave (SERW) looks promising as a way to increase the storage density. The SERW hologram offers a method other than traditional methods of selectivity, such as spatial de-correlation between recorded and reconstruction fields. In this report we present the experimental results of the SERW-hologram memory module with solid-state architecture, which is of particular interest for space operations.

TABLE OF CONTENTS

1. INTRODUCTION

New data storage technologies are required to sustain the increasing volumes and rates of data generation and to meet the demands of high throughput networks. Although two major surface storage techniques (magnetic and optical) have accelerated their roadmaps to increase the surface density and storage capacity, their technology (or cost) limits are in the 10^{11} b/cm^2 range[1,2]. Thus, alternative storage technologies, and in particular optical technologies, effective in providing a greater storage density and capacity at reasonable cost, are required [3,4,5].

Holographic memory has been a subject of interest for decades since it was first suggested by van Heerden[6]. High information density, parallel access and high-speed retrieval are among the features that make this technique of data storage so attractive. Selective properties of volume hologram due to angular[7,8] or wavelength deviation[9] from the Bragg conditions, as well as reference beam phase encoding[10] are the methods frequently used for data input and retrieval. Combination of reference beam phase encoding with spatial-shift multiplexing was shown to be an efficient approach for high-density holographic information storage.[11,12] A similar technique with many plane-wave reference (or spherical wave) beam encoding was suggested and experimentally demonstrated.[13] Although all these methods allow storing the holograms with high density, the longitudinal shift component of the recording media has not been considered for coding the individual "pages" of information. In contrast, the recently discussed method of multilayer optical storage[14,15] makes it possible to record the information at different depths within the media similar to that of a magnetic disc stack which are currently used in PC's. It was demonstrated[16] that thin layers of a holographic recording in the volume media can provide a high information storage capacity and fast data transfer rate.

We have demonstrated the possibility of building a single volume, multilayer holographic optical memory using a speckle encoded reference beam[17]. The novel features of using a spatial-encoded reference beam to multiplex holograms are that: there is high shift selectivity in any direction within the recording plane, as well as in the longitudinal direction, and phase pre-encoding can be used to build a solid-state memory system. These features lead to the possibility of building a robust, large capacity, high-data rate memory system. Moreover, this system can be built in a solid-state configuration with no moving part.

2. FUNDAMENTAL PRINCIPLES OF SERW HOLOGRAPHIC MEMORY

In proposed memory architecture the volume hologram is

formed by a plane signal wave $S_o(r) = \exp[ik_S r]$ and SERW with complex amplitude $R_o(r)$ and angular divergence $\Delta\theta_{sp}$ (see Fig. 1). For the convenience of the analysis the SERW propagates normally to the recording media front surface and the $S_o(r)$-wave has an incident angle of θ_o ($k_S = k_o\{\sin\theta_o; \cos\theta_o\}$, $k_o = 2\pi/\lambda$). The illuminated area of the diffuser is $P(q)$, ($q = \{x,y\}$), and fully developed speckle pattern[18] is formed at the distance $z_{DH} << D$. We suppose that the following relationships[19,20] between the thickness T of the hologram and recording geometry are satisfied:

$$T >> \lambda/(\Delta\theta_{sp})^2 \geq \lambda/\theta_o. \tag{1}$$

Eq. (1) describes the conditions, when the thickness of the hologram T is larger than the fringe spacing Λ, covering several speckle lengths $<\sigma_{||}>$. By fulfilling Eq. (1) the hologram reconstructs single diffraction order and provides low light-scattering level[20].
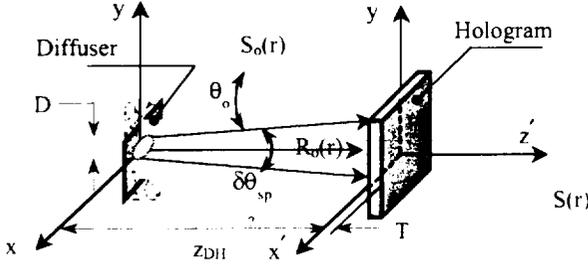
Fig. 1. SERW hologram recording geometry.

Exposure of the recording medium with the initial permittivity ε_m to the interference pattern $E_o(r) = S_o(r) + R_o(r)$ results in hologram formation that is the spatial modulation of $\tilde\varepsilon(r)$ in the following form:

$$\tilde\varepsilon(r) = \varepsilon_m + \delta\varepsilon(r) = \rho|E_o(r)|^2. \tag{2}$$

Here $\delta\varepsilon(r)$ is the modulated component of permittivity ($\delta\varepsilon(r) << \varepsilon_m$), and ρ is the recording medium parameter.

The amplitude $S(r)$ of the diffracted beam in the first Born approximation $|R(r)| >> |S(r)|$ can be found as[21,22]:

$$S(r) = -\frac{\rho k_o^2}{n_o} \int P(r)\delta\varepsilon(r)R(r)G(r,r')d^3r \tag{3}$$

where $G(r,r') = \exp\{ik_o|r-r'|\}/4\pi|r-r'|$ is the Green's function.

We now assume that after its recording, the hologram is illuminated with the same SERW, but their mutual spatial position is altered on the distance $\Delta(r)$. Such spatial shift leads to spatial decorrelation between the structure recorded in the volume of the hologram and speckle pattern in the reconstruction SERW. As the last one remains essentially the same, the reconstruction beam $R(r)$ after shift operation can be described as $R(r) = R_o(r + \Delta)$, where $\Delta = \bar{x}\Delta_x + \bar{y}\Delta_y + \bar{z}\Delta_z$ is the vector of spatial

displacement with its lateral $\Delta_\perp = \sqrt{\Delta_x^2 + \Delta_y^2}$ and longitudinal $\Delta_{||} = \Delta_z$ components.

As the magnitude of $\delta\varepsilon(r)$ in Eq. (3) is proportional to $R_o^*(r)S_o(r)$, the normalized diffraction efficiency η_N (that is the ratio of the diffraction beam intensity at its shifted state to its original value at $\Delta = 0$) can be expressed as:

$$\eta_N = \frac{I_S(r,\Delta)}{I_S(r,\Delta=0)} = \frac{\left|\int P(r')R^*(r')R_o(r')d^3r'\right|^2}{\left|\int P(r')R_o^*(r')R_o(r')d^3r'\right|^2}, \tag{4}$$

where $I_S(r,\Delta) = |S(r,\Delta)|^2$ is the diffracted beam intensity.

On the other hand, if the condition of the central-limit theorem of statistics is fulfilled and $R(r)$ is the complex Gaussian random variable at each local point of the space (x', y', z') the integration over lateral coordinates in Eq. (4) can be replaced by averaging over the ensemble, and therefore

$$\iint_S R^*(q',z_o)R_o(q,z_o)d^2q = <R^*(q',z_o)R_o(q,z_o)>. \tag{5}$$

where $q = f(x,y)$. The product on the right side of Eq. (5) is known as mutual intensity of the speckle pattern. It is equivalent to the autocorrelation function of light intensity distribution described as $C_I(r,r') = <I_R(r)I_R(r')>$, when the conditions of slowly varying average intensity in the speckle pattern $<I(r)>$ are satisfied.

We will now examine the effect of spatial decorrelation between the SERW and the structure recorded in the volume of the hologram on the efficiency of the interaction. The experimentally controlled parameter in this case is the diffracted beam intensity $I_D \sim |S|^2$ and therefore it is convenient to introduce the relative diffracted beam intensity $I_{DN}(\Delta) \sim I_D(\Delta)/I_{D(\Delta=0)}$, where the measured diffracted beam intensity $I_D(\Delta)$ is normalized by its peak value at zero lateral shift $I_{D(\Delta=0)}$. The dependence, of $I_{DN}(\Delta)$ can be expressed as[11]:

$$I_{DN}(\Delta_\perp) = \frac{\left|\int_0 e^{\frac{ik_o n\Delta_z}{2d^2}} \iint |P_D(\bar{q})|^2 \times e^{\frac{ik_r r}{d^2}q\Delta} d^2q dz\right|^2}{T^2 \times \iint |P_D(\bar{q})|^2 d^2q}. \tag{6}$$

where n is the recording medium refraction index, d is the distance between the hologram and diffuser. $P_D(q)$ is determined by the illuminated part of the diffuser.

It follows from Eq. (6) that any spatial mismatch between the hologram and reconstruction beam $R(r)$ should result in a decline of the diffracted beam intensity as plotted in Fig. 2. It shows the fall-off in $I_{DN}(\Delta_\perp)$ that occurs for lateral shift ($\Delta_{||} = $ const.) calculated from Eq. (6) in conditions that are close to experimental, i.e. $\lambda = 515$ nm,

T = 3 mm, n = 2.2 for LiNbO$_3$ crystal, d = 200 mm, P$_D$(q) = 3 mm. The dashed line in this figure comes from the calculation of the correlation function of the speckle pattern[23] for identical conditions.
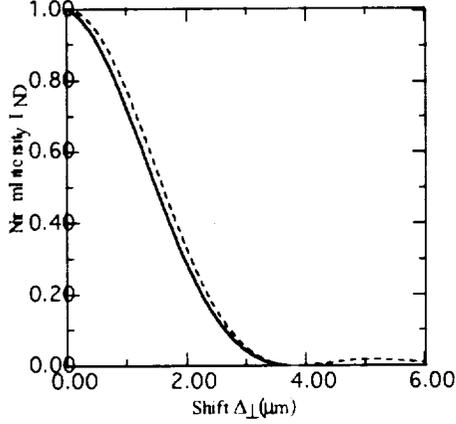

Fig. 2. Shift selectivity I$_{ND}$(Δ$_\perp$) calculated from Eq. (6)

3. MODELING THE SOLID-STATE SERW HOLOGRAPHIC MEMORY

To investigate a solid-state configuration memory, we have performed an analysis of hologram reconstruction in which the incoming beam has an orthogonal field distribution relative to the beam used in the recording step. Taking into account that the laser modes are described by Laguerre polynomials, they can serve as the simplest example of the orthogonal mode functions. Thus, we assume that at the time of the hologram recording, the laser cavity was adjusted to the TEM$_{pk}$ mode. The laser mode is switched then to TEM$_{mn}$ at the reconstruction, so that R$_o$(r) and R(r) are the wave functions of these two modes and can be described through:

$$R\ (q,\varphi)= C_{p'}E_{p'}(q)\cos(l'\varphi)\qquad (7)$$

where $E_{p'}(q)=\left(\frac{\sqrt{2}q}{\rho_R}\right)L_p^{l'}\left(\frac{2q^2}{\rho_R^2}\right)\exp\left(-\frac{q^2}{\rho_R^2}\right)$

q and φ are the polar coordinates. $C_{p'}$ is the normalizing coefficient, ρ_R is the laser bean diameter in plane of the hologram, and $L'_p(a^2)$ are the Laguerre polynomials.

The product in Eq. (5) for the field in form (7) can be transformed to:

$$R_o^*(\bar{q})R(\bar{q}) \cong \int_0^D \int_{-\pi}^{\pi} q_1 E_{p'}(q_1,\varphi)E_{mr}(q_1,\varphi)dq_1 d\varphi, \qquad (8)$$

where E_{ah} is described by (7). Substituting Eqs. (8) and (7) into Eq. (3) and assuming D >> ρ_R, the amplitude of the diffracted signal is:

$$S(\bar{q}) = S_o(\bar{r})\exp[ik_0 \sin\theta_s]C_{p'}C_{mn}H(l)\frac{\rho_R^2}{2}\times$$

$$\times \int_0^{D^2} q_1' \exp(-q_1)L'_p(q_1)L'_m(q_1)dq_1 \qquad (9)$$

where $H(l)=\begin{cases}\pi, & \text{for } l = 0; \\ 2\pi, & \text{for } l \neq 0.\end{cases}$

The integral in Eq. (9) has a non-zero value only for $p = m$. Therefore, the diffracted signal can be observed only if the reconstructing wave is completely identical to the recording one. Then, for S(q) it becomes:

$$S(\bar{q})=\begin{cases}S(\bar{q})\exp[ik_0 \cdot \sin\theta_s] & \text{for } R(\bar{q})= R_0(\bar{q}) \\ 0, & \text{for } R(\bar{q}) \neq R_0(\bar{q})\end{cases} \qquad (10)$$

It follows from the derived expressions that the amplitude of the diffracted signal is zero when the hologram is reconstructed with any wave function with its complex amplitude orthogonal to the original reference wave R$_0$(q). Thus by introducing such wave-front modulation, one can achieve multiple data storage in solid-state configuration.

4. EXPERIMENTAL SETUP

The experimental setup shown in Fig. 3 was constructed on a 4'x8' vibration isolated table. SERW holograms were recorded in a 2.8-mm-thick Fe-doped (0.02% Fe/mol) LiNbO$_3$ crystal. The crystal, with its C-axis perpendicular to the incoming recording beams, was set onto a XYZ computer-controlled positioning table, which had a precision of 0.025 μm in the X, Y, and Z planes. A CW argon laser (λ = 515nm, P = 40 mW in plane of recording media) was expanded to 1 cm and used as the coherent light source for hologram recording.
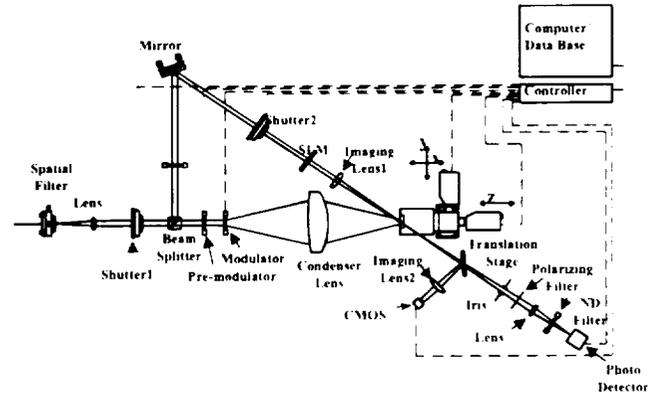

Fig. 3. Experimental setup used to record and playback the SERW holograms.

The SERW beam and the signal beam (a plane wave or an image formed with SLM), intersected at an angle of θ_0 = 35° inside the recording medium (θ_{R_0} = 0 and θ_{S_0} = 35, respectively). The average lateral speckle size at the recording media plane was calculated to be <σ> ≈ 3 μm. Diffraction efficiency of the recorded hologram in its

original position ($\Delta_\perp = 0$) could vary from 0.1 to 10^{-3}, depending upon the specific goal of the experiment.

An important component of the setup and entire memory module is the image sensor, since it determines the data transfer rate, image quality, noise, etc. The CCD cameras or CMOS sensors served this purpose. a CMOS sensor has some advantages over a traditional CCD camera as it has better pixel match with existing SLM, a more compact design, the possibility of direct pixel access for data transfer, and direct on-board windowing and processing. In our case, a CMOS VLSI sensor (VV5404) with 256 x 392 pixels is used for image capturing.

5. EXPEIMENTAL RESULTS

For the basic design of a solid-state multiplexing system, a low-resolution spatial light modulator (SLM) codes the reference beam before it strikes the speckle encoder. By changing the pattern on the SLM, an entirely new speckle pattern is generated and thus a new hologram can be addressed.

In first set of the experiments, the variation of the mode structure of the laser beam altered the spatial distribution of the complex amplitude of the beam that illuminated the speckle-encoder. This operation can be achieved by altering the laser cavity parameters. Such an operation leads to formation of the beams with orthogonal distribution of their complex amplitude. A typical example of the complex amplitude spatial distribution A for two fundamental laser modes TEM_{00} and TEM_{01} is shown in Fig. 4.
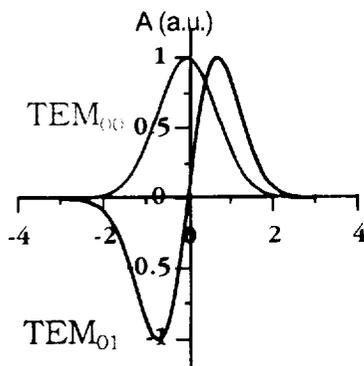

Fig. 4. Complex amplitude distribution for two fundamental laser modes

As the laser modes are described by the Laguerre polynomials, now when the hologram is recorded with for example, mode TEM_{mn} its reconstruction with any other non-equivalent mode TEM_{kl} (with its amplitude orthogonal to TEM_{mn}) should result in strong decrease of the diffracted signal intensity. The results of typical experimental realization of such operation are demonstrated in Fig. 5.
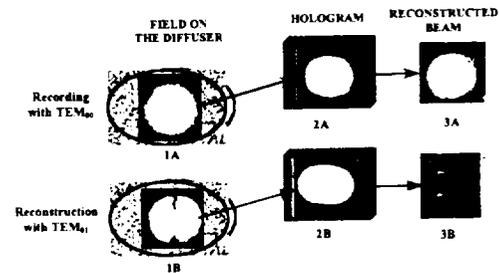

Fig. 5. Hologram recording and reconstruction with the orthogonal modes.

Here the row A corresponds to the recording (1A) and reconstruction (3A) of the hologram with the SERW formed at diffuser illumination with the laser beam having TEM_{00} mode. When the reconstruction beam with TEM_{01} was used to illuminate diffuser (1B) the diffracted signal disappears completely (3B). Thus, by changing the complex amplitude distribution from one state to another (both states being orthogonal) we are able to build the holographic data storage module with the solid-state architecture.

In another experiments we simulated the operation of a solid-state architecture by introducing both the amplitude and phase masks in front of the speckle-encoder shown in Fig. 3 and in more detail in Fig. 6.
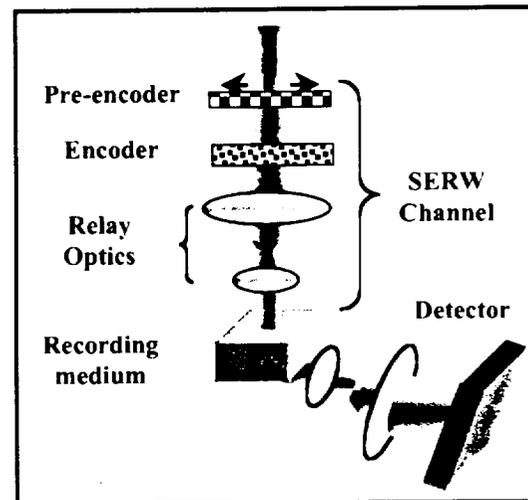

Fig. 6. Optical scheme to verify solid-state memory configuration effectiveness.

The masks had features on the order of 12 microns to simulate the pixel size of spatial light modulators. Experiments were conducted by recording a hologram with the mask in place, reconstructing the hologram, and subsequently moving the mask. Movement of the mask simulated the change that would occur when electronically-shifting the pixels of a spatial light modulator or beam steering over the speckle generator. It was found that the reconstructed intensity decreased as the mask was moved. A typical example of such pre-encoder in action is shown in Fig. 7a, where its spatial shift resulted in a smooth decrease of the diffracted signal
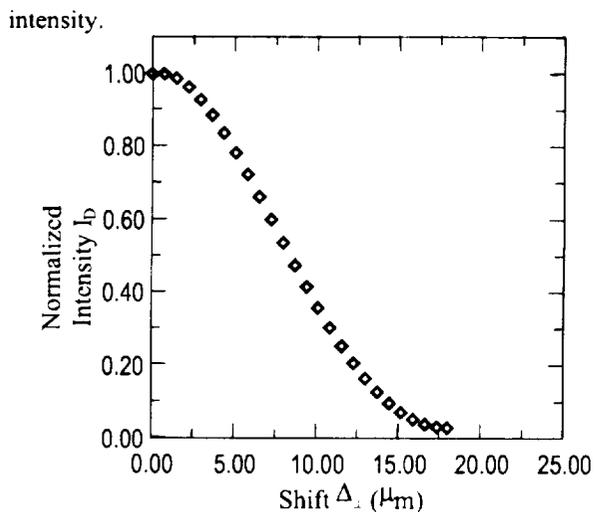
intensity.

a)
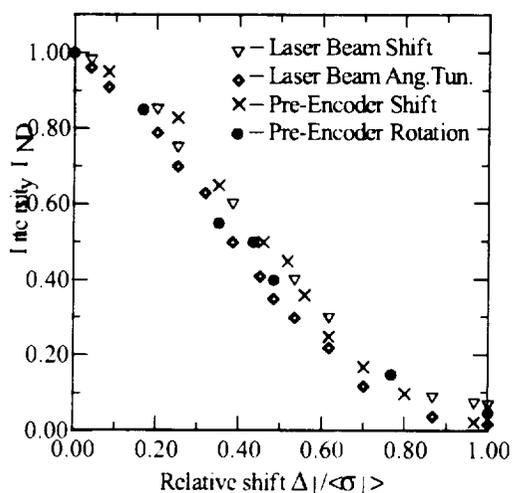
b)

Fig. 7. Dependence of the diffracted signal intensity upon the spatial shift of the pre-modulator (a) and the results of the comparison of several pre-modulation techniques (b).

The phase mask had much greater selectivity than the amplitude mask and therefore a phase-only SLM should provide the best results. Identical data have been received at an equivalent spatial decorrelation in any direction in the XY plane, similar to that of physical shift selectivity. We also explored several other options in simulating the solid-state operation, applying spatial and angular deviation of the laser beam before it hits the encoder or pre-encoder, as well as pre-encoder axial revolution. The results of the performed measurements are shown in Fig. 7b, and confirm the identity in the operation's effectiveness.

6. CAPACITY OF SERW-HOLOGRAPHIC MEMORY

It is important to estimate the information density

available through the variety of combinations of 3-D-shift, angular, and wavelength selectivity. The storage capacity of the 3-D hologram with a volume where $V = d^2 \times t$ (and where d is the lateral size of the holographic media and t is its thickness) can be calculated as

$$N_V = \frac{4\pi}{3} \frac{V}{\lambda^3}. \tag{11}$$

The expression to determine the total number of multiplexed holograms that can be stored using the shift selectivity scheme is

$$N_A \approx \frac{d^2}{\Delta_\perp^2} \times \frac{t}{\Delta_\parallel}. \tag{11}$$

Finally, the correlation function of the speckle pattern is described through the coherence function of the illuminating light. So, changing the wavelength of the recording illumination enables the storage of more "layers" of the holograms and opens an additional mechanism for increasing the capacity of the memory. It was proved experimentally[2] that a spectral shift of $\delta\lambda \approx 1.5$ nm of the reconstructing wavelength results in total degradation of the diffracted signal. Thus, using a tunable laser with spectral a interval of $\Delta\lambda >> \delta\lambda$, additional "layers" of information can be stored. The limiting value for information density to be recorded using speckle-encoded reference beam can be found as

$$N_{max} = N_A \times \left(\frac{\Delta\lambda}{\delta\lambda} \right). \tag{12}$$

This leads to the data storage capacity for $1 \times 1 \times 1$ cm^3 recording media using a speckle reference beam with the following characteristics: $\Delta_\perp \approx 3$ µm; $\Delta_\parallel \approx 30$ µm; and $\delta\lambda/\delta\lambda \approx 100$: $N_{max} \approx 3 \times 10^{11}$ holograms. If every hologram is a page of information with 1 Mbit, then total capacity can be estimated as 3×10^{17} bits; however, taking into account the limited range of the recording media refraction index modulation, the most optimal combination of the selectivity will be chosen which optimizes the number of holograms that can be stored. This selection will be focused towards development of a solid-state memory system. The dynamic range for the recording media refraction index modulation is limited and is normally within the frame of $\Delta n \_ 1.3\_10^{-2}$ for photopolymer materials and $\Delta n \_ 10^{-4}$ for photorefractive crystals. Therefore, the proper and most efficient choice of the selectivity mechanism to be used is particularly important, as it determines the storage density and capacity of the memory unit.

With optimized CMOS detectors, the average diffraction efficiency can ultimately be around 10^{-8}, which corresponds to a refraction index modulation on the level of $\Delta n \_ 1.6 \times 10^{-8}$. With this value of diffraction efficiency for each individual hologram, about 10^5 holograms can be multiplexed in a PRC and 10^6 in photopolymer (100 Gbit and 1 Tbit, respectively) at one location. To realize such a high storage capacity, a thorough study of multiplexing is of primary importance.

110

7. CONCLUSION

In conclusion we demonstrated the possibility of building high-density holographic memory module with solid-state architecture by using the properties of the volume hologram recorded and reconstructed with spatially encoded reference beam. This memory module can operate with both volatile and non-volatile recording media. -

8. ACKNOWLEDGEMENTS

9. REFERENCES

[1] M. Keshner, Status of optical storage in Japan, in WTEC panel report on the future of data storage technologies, June 1999 (National Science Foundation International technology Research Institute).

[2] The report on sub-terabit optical memory project can be found at http://www.oida.or.jp

[3] NSIC/OIDA Optical disk storage roadmap, National storage industry consortium, pp. 3-15, 1997

[4] C. Wilde, Easier Archiving, Information week, August 23, 1999.

[5] A. A. Jamberdino, "Storage and retrieval: the next generation" Report at the SPIE-3208 conference, "Advanced optical data storage: Materials, systems, and interfaces to computrs", July 1999.

[6] P.J. Van Heerden, Theory of optical information storage in solids, Appl. Opt. 2, 156 (1963).

[7] F.H. Mok, Angle-multiplexed storage of 5000 holograms in lithium niobate, Opt. Lett. 18, 915 (1993).

[8] S. Tao, D. Selvian, J. Midwinter, Spatioangular multiplexed storage of 750 holograms in an Fe:LiNbO$_3$ crystal, Opt. Lett. 18, 912 (1993).

[9] J. Rosen, Dynamic phase-encoding storage of 64 images in a BaTiO$_3$ photorefractive crystal, Opt. Lett. 18, 744 (1993).

[10] Rakuljic, G.A.; Leyva, V.; Yariv, A. Optical data storage by using orthogonal wavelength-multiplexed volume holograms. Opt. Lett. 17, 1471 (1992).

[11] A. Darskii, V. Markov, Some properties of 3D holograms with a reference speckle-wave and their application to information storage, SPIE Proc. V.1600, pp. 318-332, 1992

[12] V. Markov Yu. Denisyuk. R. Amezquita, "3-D Speckle-shift hologram and its storage capacity", Optical Memory and Neural Networks v. 6, No 2. pp. 91-98, 1997.

[13] D. Psaltis, M. Leven, A. Pu et al., Holographic storage using shift multiplexing, Opt. Lett. 20, 782 (1995).

[14] J. Rosen, Multilayer optical recording (MORE), SPIE Proc. 2514, 14 (1995).

[15] S. Chinn and E. Swanson, Multilayer optical storage by low-coherence reflectometry, Opt. Lett. 21, 899 (1996).

[16] P.A. Mitkas, L.J. Irakliotis, G. Betzos, "Optical Processing Paradigm for Electronic Computers", IEEE Computer, 31, 2, pp. 45-51, 1998.

[17] V. Markov, J. Millerd, J. Trolinger, M. Norrie, J. Downie, D. Timuchin, Multiyear holographic optical memory, Opt. Lett., v. 24, n° 4, pp. 265-267, 1999.

[18] E. Jakeman, Enhanced backscattering through a deep random phase screen, J. Opt. Soc. Am. A 5, 1638-1648, 1988.

[19] V. Sidorovich, Diffraction efficiency of three-dimensional phase holograms, Sov. Phys. Tech. Phys., v. 21, pp. 742-745, 1976.

[20] B. Zel'dovich, V. Shkunov, T. Yakovleva, Holograms of speckle field, Soviet Physics - Uspekhi, v.29 pp. 678-702, 1986.

[21] V. Markov, M. Soskin, A. Khizniyak, V. Shishkov, Transformation of the structure of coherent light beams by volume phase holograms in LiNbO$_3$, Soviet Techn. Phys. Lett., v.4, pp.304-6. 1978

[22] A. Darskii, V. Markov, Opt. Spectrosc. Shift selectivity of the holograms with speckle reference wave v. 65, pp.392-394, 1988

[23] J. W. Goodman, "Statistical properties of laser speckle pattern" in book "Laser speckle and related phenomena" Ed. J. Dainty, Springer-Verlag, 1984.

BIOGRAPHY

Dr. Vladimir Markov received his Ph.D. in physics and mathematics from the Institute of Physics of the National Academy of Sciences of Ukraine and MS degree in radiophysics from the Kiev State University (Ukraine). From 1978 he was the Head Center of Applied Holography at the Institute of Physics of the Academy of Sciences of the Ukraine. In 1994 he established, and became a director of the Institute of Applied Optics of the National Academy of Sciences of Ukraine. In 1998 he joined MetroLaser Inc., where he is the head of the Applied Optics Group. He is a member of SPIE, OSA and AAAS (American Association for the Advancement of Sciences), member of the board of SPIE-Ukraine, Optical Society of Ukraine, and a member of the advisory Editorial Board of the journal Optics & Laser Technology.

SESSION #4: New Technologies

Thursday, November 16, 2000
9:20 AM – 12:00 PM
Chair: *Scott Tyson,* Air Force Research Laboratory,
Albuquerque, New Mexico

9:20 AM   Chalcogenide-Based Non-Volatile Memory; *J. Maimon and E. Spall,*
            Ovonyx Inc, *R. Quinn and S. Schnur,* Lockheed Martin Federal Systems

9:40 AM   Advances in FeRAM Technology; *G. Derbenwick, D. Kamp, S. Philpy*
            and A. Isaacson, Celis Semiconductor

10:00 AM   High Density Packaging of Non-Volatile Memory Systems; *J. Carson*
            and Keith Gann, Irvine Sensors

10:20 AM   (Break)

10:40 AM   All-Metal Magnetic RAM; *E. Torok and R. Spitzer,* IME

11:20 AM   Advanced MRAM Concepts; *J. Daughton,* Nonvolatile Electronics

Chalcogenide-Based Non-Volatile Memory Technology

J. Maimon (jmaimon@ovonyx.com), E. Spall (espall@ovonyx.com)
Ovonyx, Inc., 8665 Sudley Rd., #263, Manassas, VA 20110

R. Quinn (robert.quinn@lmco.com), S. Schnur (steven.schnur@lmco.com)
Lockheed Martin Federal Systems, 9500 Godwin Drive, Manassas, VA 20110

Abstract— Chalcogenide is a proven phase change material used in re-writeable CDs and DVDs. This material changes phases, reversibly and quickly, between an amorphous state that is dull in appearance and electrically high in resistance, and a polycrystalline state that is highly reflective and low in resistance. The application of this commercially proven technology to create dense, high-speed, non-volatile semiconductor memories is discussed.

TABLE OF CONTENTS

1. INTRODUCTION

Historically semiconductor memory technology has developed in such a way that currently available memory components can be divided into categories representing their specialized features, and no one single technology combines all the desirable memory characteristics of highest speed, lowest cost, lowest power, long cycle and non-volatility. Thus we have DRAM, SRAM and FLASH as the dominant basic memory technology types providing the spread of characteristics given in Table 1.

Table 1. Memory Technology

Mem Type	Speed	Power	Cost	Cycle Life	Non-Volatile
SRAM	Very High	High	High	Very High	No
DRAM	High	Med.	Very Low	Very High	No
FLASH	Low	Very Low	Low	Low	Yes
OUM (Potential)	High to Very High	Low to Very Low	Very Low	High to Very High	Yes

An ideal semiconductor memory technology would combine or "Unify" the attractive features of these technologies without acquiring any of the unattractive features. Such a memory technology, Ovonic Unified Memory (OUM) is now being developed using the class of elements known as chalcogenides. It is expected that this technology will eventually allow chips that have SRAM speed, DRAM cost and power characteristics and non-volatility as shown in Table 1.

The importance of non-volatility in the marketplace is key as whole classes of new high volume portable products are now emerging that require non-volatility such as hand-held cell phones. Personal Digital Assistants (PDAs), electronic digital cameras and others. Additionally, a dense, nonvolatile memory that is radiation resistant as well as low in both cost and power is one of the most pressing needs for electronics operating in the space environment. It is also interesting to contemplate the total amount of time that would be saved by eliminating the boot up time required by all personal computers using DRAM.

The storage medium for the OUM chips is the same alloy currently being used in large-scale production to produce re-writeable CD disks [1]. This chalcogenide based memory technology holds the promise of providing almost ideal characteristics at current lithography levels and allowing scaling down through the 0.1μm regions because the performance characteristics improve as the dimensions shrink. Ovonyx Inc. is currently developing this technology under license from Energy Conversion Devices Inc., the developer of the base technology.

2. CHALCOGENIDE TECHNOLOGY: CD-RW AND DVD-RAM

The term "chalcogenide" refers to the Group VI elements of the periodic table. It has also been used to refer to alloys containing at least one Group-VI element such as the alloy of germanium, antimony, and tellurium discussed here. Energy Conversion Devices, Inc. has used this particular alloy to develop a phase-change memory technology [2]-[7] used in commercially available re-writeable CD and DVD disks. This phase-

change technology uses a thermally activated, rapid, reversible change in the structure of the alloy to store data. Since the binary information is represented by two different phases of material it is inherently non-volatile, requiring no energy to keep the material in either of its two stable structural states.
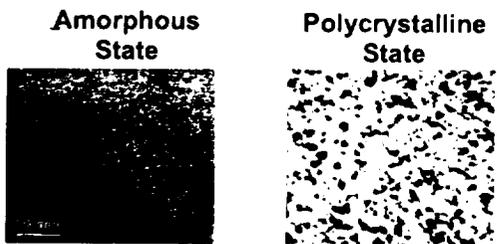

Fig. 1. Transmission Electron Microscope images of the two phases of a GeSbTe alloy

The two structural states of the chalcogenide alloy (as shown in Figure 1 are an amorphous state (no long-range order of atoms) and a polycrystalline state (composed of many crystals, each having atoms placed in a repetitive order). Relative to the amorphous state, the polycrystalline state shows a dramatic increase in free electron density (similar to a metal). This difference in free electron density gives rise to a difference in reflectivity. In the case of the rewriteable CD and DVD disk technology, this difference in reflectivity is used to read the state of each memory bit by directing a low-power laser at the material and detecting the amount of light reflected.

Programming data (by changing the phase of the material) is accomplished by heating the material to an elevated temperature. Depending upon the temperature achieved the material is either melted to form the amorphous state or crystallized by holding it at a lower temperature for a slightly longer period of time (Fig. 2). For the CD and DVD technology, the heat is generated from a high power laser.
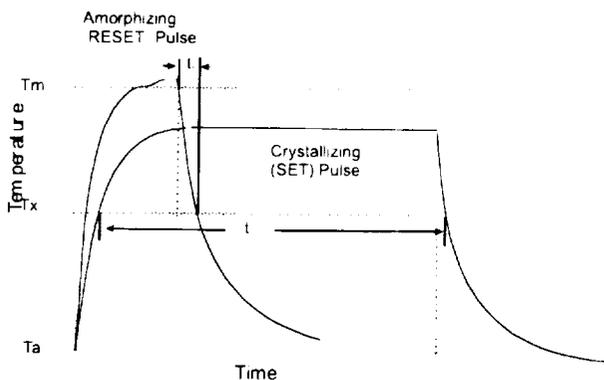

Fig. 2. Chalcogenide programming

3. CHALCOGENIDE TECHNOLOGY: SEMICONDUCTOR MEMORIES

The two phases of the chalcogenide alloy have important differences in electrical properties due to the change in free electron density. The resistivity of the polycrystalline state has been shown to be up to four orders of magnitude lower than that of the amorphous state. Another interesting electrical property of the material is that, above a certain applied voltage (V_T), the two states "switch" to a low resistance state that is independent of the material phase (Fig. 3). This switching does not change the structural phase of the material, and the resistance values can be restored once the applied voltage is brought below the "holding" voltage (V_H). This phenomenon allows for high current values to be forced through the device with a relatively low voltage that is independent of the phase of the material. These electrical properties are very attractive when considering using a chalcogenide alloy in a semiconductor memory device. The state of such a device can be read non-destructively by charging a capacitively-loaded line through a variable resistance chalcogenide element held at a voltage below V_T [8]. The rate of charging — a function of the chalcogenide resistance — can be detected and converted into a logic state. Programming the chalcogenide element is achieved by passing a sufficiently high current (with an applied voltage above V_T) to heat the material to the crystallization or melting temperature.
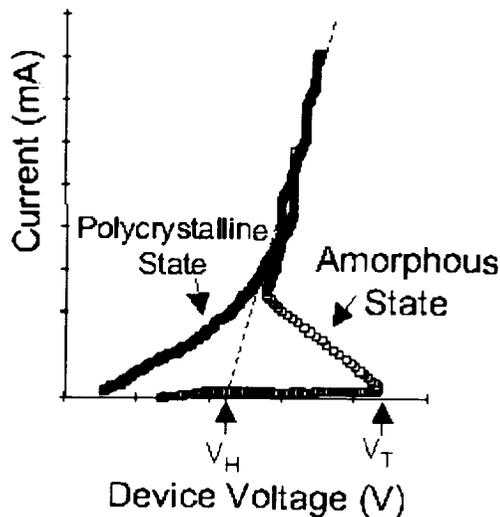

Fig. 3. I-V characteristics of chalcogenide

Electrically addressable chalcogenide memory elements have been built to show proof of concept of a chalcogenide-based semiconductor memory. Figure 4 demonstrates the ability to electrically change the resistance of the chalcogenide element by passing a

current through the device. For low currents the device resistance is independent of current and is only a function of the initial state of the material. For moderate currents the temperature of the device is high enough to crystallize the amorphous material, causing its resistance to drop. For higher currents the temperature is high enough to melt the polycrystalline material, increasing the resistance. It is important to realize that for each point on this graph, the resistance is not measured at the indicated programming current, but rather the programming current is applied and then the device resistance is measured at a low current value (0.1mA).

Unlike flash non-volatile memories, where long erase sequences are needed prior to writing data, chalcogenide elements can be directly over-written without the need for first erasing the previous stored state. Programming of these chalcogenide memory elements has been achieved with pulse widths as small as 3ns.
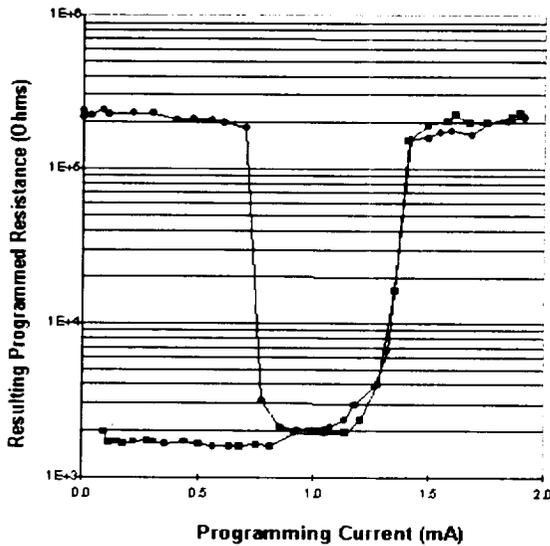
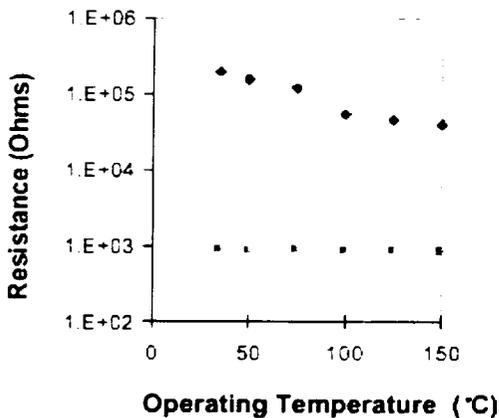Fig. 4 Electrical programming of chalcogenide

Fig. 5 Chalcogenide resistance versus temperature

Figure 5 shows that at temperatures as high as 150°C a resistance spread of at least 40x can be maintained for the two phases of the material. This spread increases at lower temperatures. Figure 6 shows that it is projected that data stored in chalcogenide resistors could be retained for 10 years of continuous operation at approximately 125°C. Data retention at typical operating temperatures would be considerably longer.
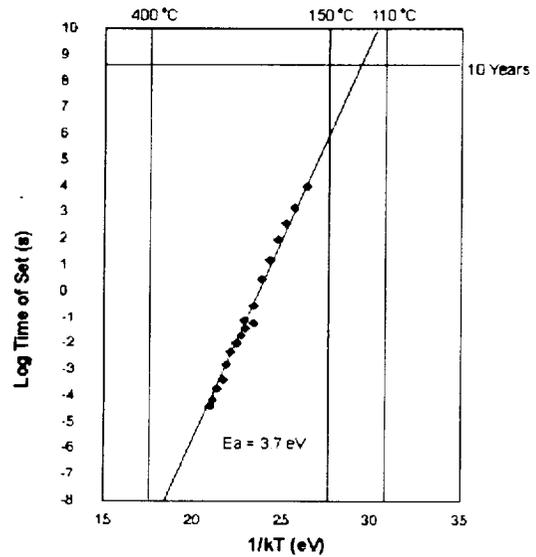

Fig. 6 Data retention given as an Ahrennius time-temperature relationship

Because of their proliferation in today's powerful computational systems, the cost of semiconductor memories is an important consideration. For chalcogenide-based memories the manufacturing cost should be very low. As the volume of the chalcogenide element decreases, the programming current (power) is reduced and the memory element cell operates more quickly. Chalcogenide-based memory cells have been designed that would be smaller than DRAM cells. Fabrication of the chalcogenide element itself requires between two and four masking steps in a planar topology. These steps are performed after the support or logic transistors have been completed and do not require any process changes that would affect the performance of the transistors (unlike DRAM capacitor processing).

Chalcogenide memory elements also show the promise of a tremendous decrease in the cell size (and cost) per bit by allowing true storage of multiple bits per memory element. Figure 7 shows that the transition between low and high resistance can be reproducibly achieved in several steps instead of in one abrupt transition. The 16 discrete resistance states shown could store 4 bits in one small chalcogenide element.
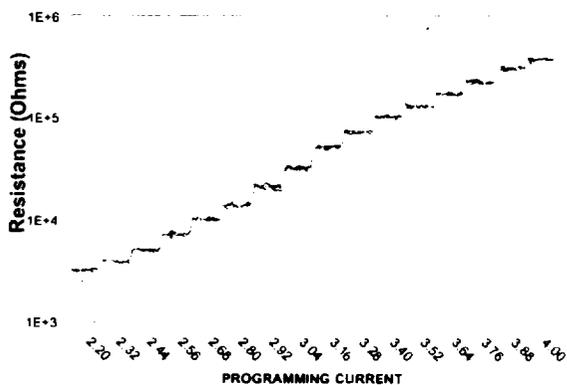
Fig 7. Multiple state storage example showing 16 separate states. For this graph, 10 pulses were performed at each of 16 current/resistance steps, and the overall measurement was repeated 10 times (each state represents 100 separate write/reads).

4. CURRENT PROGRAM DESCRIPTION

Energy Conversion Devices (ECD) developed the data described to this point during the late 1990's. This work concentrated on devices in the 1μm device size range and generally yielded switching currents in the 20mA region. Devices with lower switching currents were made using e-beam lithography for one photo level, but the devices were still overly large in size. In order to be able to make high-density memory chips using this technology it was necessary to change the structure of the device to significantly reduce the overall area, and to reduce the switching currents to below 1mA.

Ovonyx Inc. is applying the OUM technology to semiconductor memories under license from ECD. Ovonyx currently has two joint development programs underway, one with Intel Inc. and one with Lockheed Martin Corporation at its Space Electronics & Communications (LMSEC) site in Manassas, Virginia. The following describes the LMSEC activity that is being funded by the Space Vehicles Directorate of the Air Force Research Laboratory.

The Chalcogenide-based Random Access Memory (C-RAM) program began in November of 1999 with the goal of integrating the phase-change material into a radiation-hardened digital CMOS fabrication process. The program aims at reducing the switching currents required to change device state and then integrating the results into a memory chip of significant density.

The program will first develop a memory cell that exhibits switching at under 1 mA in 20ns or less, and then integrate that cell into a high density memory chip using LMSEC's radiation hardened RHCMOS process. The first demonstration will be a 16Mb chip using 0.5μm lithography by the end of 2002.
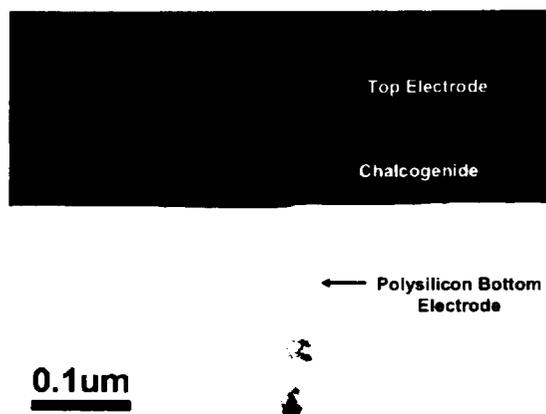

Fig. 8. Transmission electron microscope image of functional chalcogenide storage element fabricated under the C-RAM program
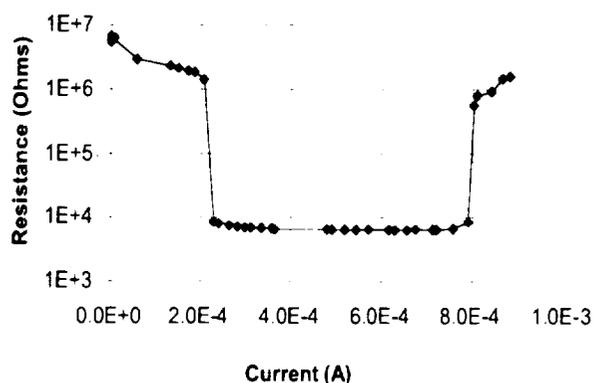

Fig. 9. Typical R-I plot for C-RAM devices

5. RESULTS TO DATE

Experiments conducted at Lockheed Martin's facility and Ovonyx Inc. facilities in Troy Michigan have shown the feasibility of making C-RAM storage cells with the extremely small size and very low programming currents necessary to construct a memory chip of high density. Storage cells have been constructed and tested using a variety of electrode compositions and physical arrangements of the structures involved, with work continuing at the present time related to optimization of the materials and physical structures as well as gathering reproducibility and cycle life data.

Devices exhibiting switching speeds under 50ns with less than 1mA of switching current have been successfully fabricated and tested. These devices typically exhibit two to three orders of magnitude difference in resistance between the set and reset states, and cycle life of up to 10^{11}. A typical device cross section is shown in Figure 8 and typical programming characteristics and I/V curves are graphed in Figures 9 and 10. Figure 11 shows

representative life cycle data. Experiments continue, centered on reducing the variation of the switching current, optimizing cycle life and collecting statistically significant design data to permit design of the 16Mb chip.
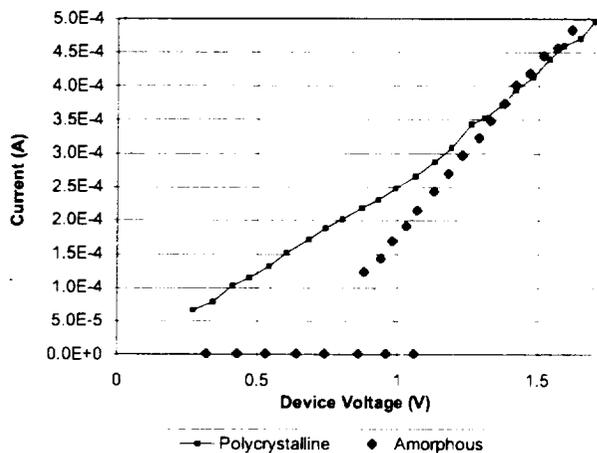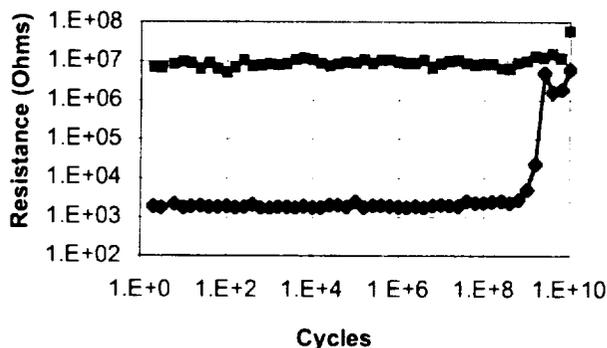

Fig. 10. Typical I-V curve for C-RAM devices

Cycles

Fig. 11. Typical cycle life for C-RAM devices

6. FUTURE PLANS

The schedule for development work leading up to the fabrication and test of a 16Mb C-RAM chip as envisioned under the C-RAM program are given in Table 2, assuming full funding of the various phases of the work including integration and design activities. The process into which the chalcogenide cell will be integrated is the 0.5μm LMSEC RHCMOS process that has been QML qualified at the 5V and 3.3V power supply levels. Both of these process variants are radiation hardened and no degradation of that hardness level is expected as a result of the combining of the technologies. Radiation testing of chalcogenide storage devices in the total dose environment has been reported by Bernacki et al [9]. They reported no observable effects on stored data up to

total dose levels of 2Mrad, and are planning to follow up with neutron testing of devices later this year.

TABLE 2. C-RAM Development Schedule

Memory Element Defined	12/2000
Integration with RHCMOS	12/2001
16M C-RAM Demonstration	6/2002
Qualification of 16M C-RAM	12/2002

Development will proceed following the definition of the C-RAM memory element with design of a Comprehensive Technology Characterization Vehicle (CTCV) which will be a test site allowing for the measurement of electrical and other parameters. The CTCV will be useful for demonstration of several drive/sense schemes, which will allow a staged development of the final 16M design. The CTCV will also serve as the vehicle to demonstrate successful integration of the chalcogenide module with the RHCMOS process and will be suitable for radiation testing of the integrated technology in the various radiation environments of interest.

ACKNOWLEDGMENT

REFERENCES

[1] S R Ovshinsky, "Phase change for optical storage," Business and Technical News from Balgers Materials, Issue 9, October 1999

[2] S. R. Ovshinsky and H Fritzsche, "Amorphous semiconductors for switching, memory, and imaging application," IEEE Trans Elect Dev vol ED-20, no 2, pp. 91-105, Feb 1973

[3] S Ovshinsky, "Reversible electrical switching phenomena in disordered structures," Phys Rev Lett vol 21, no 20, pp 1450-1453, Nov 11, 1968

[4] H Fritzsche, "Electronic phenomena in amorphous semiconductors", Annual Review of Materials Science, v2, pp 697-744, 1972

[5] S R Ovshinsky, S J Hudgens, W Czubatyj, D A Strand, and G C Wicker, "Electrically erasable memory elements having improved set resistance stability," US Patent # 5,414,271, May 1995

[6] Hernandez, J., Chao, B., Strand, D., Ovshinsky, S., Pawlik, D., Gasiorowski, P., "Crystallization Studies of Ge:Sb:Te Optical Memory Materials" Appl. Phys Comm. V11, #4, pp557-581, 1992

[7] Yamada, N., Ohno, E., Nishiuchi, K., Akahira, N.,Takao, M., "Rapid-Phase Transitions of GeTe-Sb2Te3 Pseudobinary Amorphous Thin Films for an Optical Disk Memory", J Appl Phys., v69 #5, pp2849-2857, 1991

[8] S Tyson, G Wicker, T Lowrey, S Hudgens and K Hunt, "Nonvolatile, high density, high performance phase change

memory", *Proceedings 2000 IEEE Aerospace Conf.*, Big Sky MT, Mar. 18-25 2000.

[9] S. Bernacki, K. Hunt, S. Tyson, S. Hudgens, B. Pashmakov, and W. Czubatyj, "Total dose radiation response and high Temperature imprint characteristics of chalcogenide based RAM resistor elements" *IEEE Trans. Nucl. Sci.*, vol. 47, no. 6, Dec. 2000

Jonathan D. Maimon received the B.S. degree in chemistry from The College of William and Mary in 1983 and the M.S. degree in chemical engineering from Purdue University in 1985.

From 1985 to 2000 he worked as a manufacturing engineer and then as a process development engineer developing radiation-hardened CMOS processes for IBM's Federal Systems Division, which transferred during this period to Loral and then Lockheed Martin. During this time frame he developed processes for radiation-hardened FPGAs and 64K, 256K, 1M, and 4M SRAMs. Since March 2000 he has been working with Ovonyx as Director of Process Development.

Edward J. Spall received the BSEE degree from the University of Detroit in 1963, and has worked in the semiconductor electronics industry for over 35 years beginning with 31 years at IBM where he held various design and managerial positions with the Federal Systems Division of IBM. While at IBM he designed circuitry for ICBM guidance computers and initiated and managed the semiconductor design and process fab activities at IBM Manassas for 18 years. Mr Spall oversaw the development of CMOS 64K, 256K and 1M SRAM and 16 & 32 bit Microprocessor devices which were radiation hardened to natural and strategic radiation levels and which now operate in a host of NASA, Military and commercial satellites. Mr Spall was Director of Technology at Loral and Lockheed Martin after successive sales of the business to those companies, retiring from Lockheed Martin in November 1999.

Mr. Spall then joined Ovonyx Inc. as Director of Manassas Operations involving development of the OUM technology in conjunction with Lockheed Martin.

Robert M. Quinn received the B.S., M.S., and Ph.D. degrees in electrical engineering from Rensselaer Polytechnic Institute in 1963, 1965, and 1968, respectively.

Dr. Quinn has worked in commercial silicon semiconductor process development and manufacturing for over 30 years, primarily at IBM Technology Development in East Fishkill, NY, and Essex Junction, VT. His experience ranges from 72Kbit, 5um DRAM through 64Mbit, 0.25um DRAM, with logic and SRAM derivatives from those programs beginning in the 2um silicon gate era. During his employment, he has held various Senior Engineering, Staff, and Managerial positions in development and manufacturing. Currently, Dr. Quinn works at Lockheed Martin Corporation in Manassas, VA, in Advanced Digital Solutions, where he is the Lead Technologist for their radiation hardened CMOS semiconductor development and manufacturing line.

Steven G. Schnur received the BS in Physics degree from the University of Central Florida in 1970.

Mr Schnur held process engineering positions at Martin Marietta in Orlando Florida From 1970 to 1974 and at Westinghouse in Lithicum Md from 1974 through 1980 where he was responsible for lithographic process engineering of CCD, Bipolar and CMOS processes. Mr Schnur then moved to IBM in Manassas Virginia where he progressed from process engineering through management assignments in the company's VLSI rad hard CMOS development and production line where he worked on 64K, and 256K SRAM development and manufacturing. In 1990 Mr Schnur was named program manager of the X-Ray lithography program at IBM and was responsible for a $107M development effort which resulted in the production of deep submicron X-ray lithography which was demonstrated at IBM's Advanced X-ray facility at Fishkill N.Y. Since 1999 Mr Schnur has been Manager of process development in the Lockheed Martin Space Electronics & Communications Silicon Technology Center at Manassas Virginia which produces memory, microprocessor, ASIC, FPGA and related electronic devices for the military and civilian space market.

Advances in FeRAM Technologies

Gary F. Derbenwick, David A. Kamp, Stephen C. Philpy and Alan F. Isaacson

Celis Semiconductor Corporation
5475 Mark Dabling Blvd., Suite 102
Colorado Springs, CO 80918
(719) 260-9133; (719) 593-8540 (fax); celis@celis-semi.com

Abstract—Commercial ferroelectric random access memories (FeRAMs) are emerging into the marketplace. Reliability levels have been measured to be comparable to or better than those of other reprogrammable nonvolatile semiconductor technologies. Both SBT and PZT ferroelectric materials have been used in these memories for the memory storage elements. By combining ferroelectric memory technology with radiation hardened CMOS technology, nonvolatile semiconductor memories for space applications can be attained. These memories have high endurance, fast write times and low power performance. While significant efforts are underway at major semiconductor manufacturers worldwide to develop ferroelectric memory businesses, funding limitations must be overcome to allow the development of radiation hardened ferroelectric memories for space applications.

TABLE OF CONTENTS

1. INTRODUCTION

Ferroelectric semiconductor memories are well suited for space applications because of their low power performance, nonvolatility, fast programming times, and high endurance capability. The ferroelectric storage capacitor itself is radiation tolerant. When combined with radiation hardened CMOS circuitry, FeRAMs can survive radiation doses typically encountered in earth orbit and in near and deep space missions, including ionizing radiation, proton irradiation and high energy ion bombardment.

Both PZT (lead zirconium titanate) and SBT (strontium bismuth tantalate) can be used for the ferroelectric mat-erial, depending on the product specifications. Read/write performance is comparable to that of a DRAM with the added feature that the FeRAM provides nonvolatility. At present, major semiconductor manufacturers are developing commercial FeRAMs. Space level FeRAMs are expected to leverage off the commercial development activities.

2. FcRAM DENSITY

Commercial FeRAMs have been fabricated at densities of 1M and below and most FeRAMs have used single level metal processes. For FeRAMs to compete with Flash memory, higher densities are required. Fig. 1 shows the cross section of an FeRAM fabricated with two levels of metalization and a titanium nitride local interconnect level. With the capability to fabricate FeRAMs with multi-level metal, multi-Megabit FeRAMs are closer to reality.

Figure 1. Cross Section of an FeRAM Memory Array Using Multi-level Metal Interconnect. *Figure courtesy of Hyundai Electronics Co., Ltd.*

2. FcRAM RELIABILITY

Reliability has been reported for initial commercial FeRAMs.[1,2] The FIT failure rate of a commercial 1K embedded FeRAM is shown in Fig. 2. The primary failure mechanism is a freak bit retention failure for ten year retention. Failure rates under 100 FITs can be obtained

for operation of this device at 70 °C. One FIT is one device failure per 10^9 device operating hours. Failure rates for higher operating temperatures can be determined from Fig. 2.

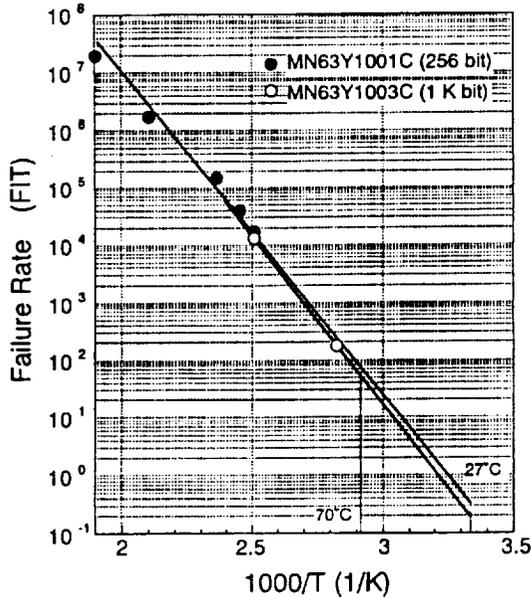More recently, excellent reliability results have been presented for a 64K FeRAM using multi-level metal.[2]

Figure 2. Failure rate of a 1K embedded FeRAM.

3. SBT and PZT FERROELECTRIC MATERIALS

Retention and endurance performance characteristics are key attributes for nonvolatile memories used in space applications, especially for deep space missions. Retention and endurance performance of FeRAMs depend on the materials used for the ferroelectric storage element and the electrodes of the ferroelectric capacitors. FeRAMs currently being commercially manufactured use PZT (lead zirconium titanate) or SBT (strontium bismuth tantalate) for the ferroelectric storage element material, depending on the product specifications.

SBT storage cells using platinum electrodes program at voltages below 3V and have been shown to have virtually unlimited endurance by testing to greater than 8 X 10^{12} write/erase cycles, as shown in Fig. 3.[2]

For PZT, capacitors formed with platinum electrodes do not provide the nonvolatile characteristics of SBT with platinum electrodes. The endurance characteristics for PZT with platinum electrodes is typically not more than 10^8 programming cycles. However, if IrO_x (iridium

oxide) or LSCO (lanthanum strontium cobalt oxide) are used as electrodes for PZT capacitors, performance similar to SBT with platinum electrodes may be obtained. As shown in Fig. 4 for a PZT storage capacitor, PZT performance with LSCO electrodes also can provide virtually unlimited endurance.
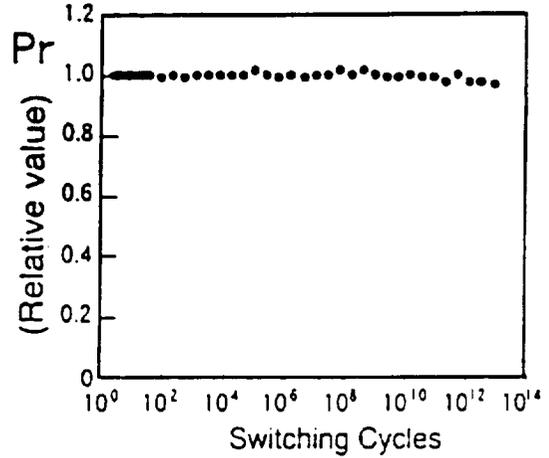

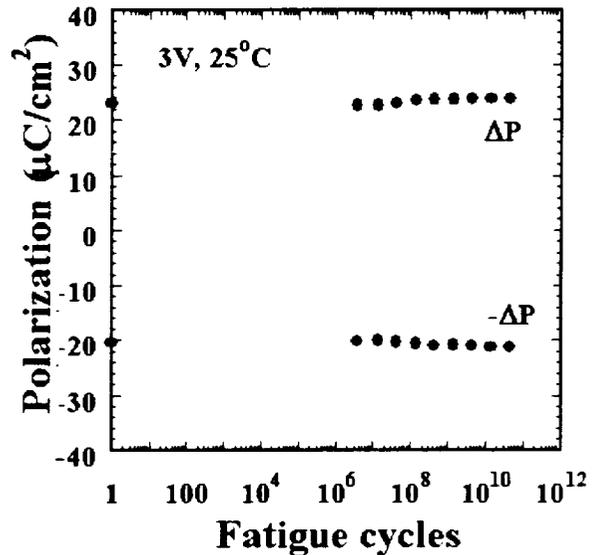Figure 3. SBT Endurance Performance normalized to the initial ferroelectric polarization, P_r.

Figure 4. PZT Endurance Performance of a PZT storage capacitor with LSCO electrodes. Data are shown for room temperature. Endurance data at 100 °C are similar. *Data courtesy of Radiant Technologies.*

Retention results have been presented for a commercial 64K FeRAM using multi-level metalization.[2] Ten year retention at 150 °C is possible, as shown in Fig. 5.

Since today's FeRAM's use a destructive read with an automatic rewrite during the read cycle, simply reading

the memory provides a full refresh for retention purposes. Therefore, ten year retention is not required for most space applications. However, the endurance specification also applies to the number of read cycles because of the destructive read operation.
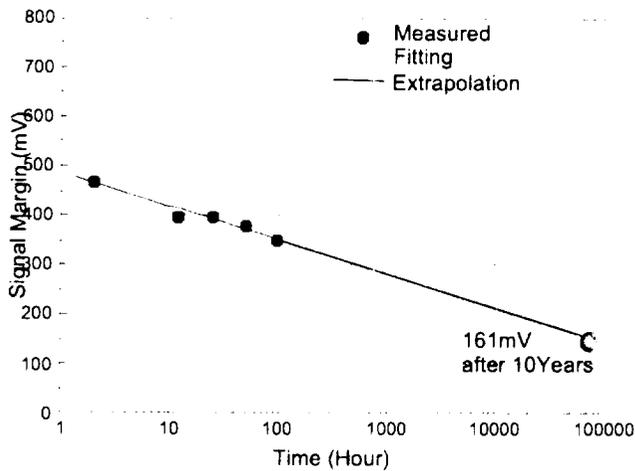

Figure 5. Retention Characteristics at 150 °C of an SBT FeRAM using multi-level metalization. *Data courtesy of Hyundai Electronics Industries Co., Ltd.*

4. FeRAM PERFORMANCE

Commercial FeRAMs are being developed that use one transistor, one capacitor memory cells similar to those used in DRAMs. However, for space applications, a more rugged architecture using a pair of memory cells with true and complement bit lines is recommended. A radiation hardened sense amplifier design is used to avoid single event upset.

Programming times for commercial FeRAMs are similar to those of DRAMs, under 100 ns. Programming times for FeRAMs designed for space applications are projected to be a few hundred nanoseconds. Since the ferroelectric programming time is longer at lower temperatures, longer programming times are required because of the extended low temperature range for space applications (-55 °C) as compared to that for commercial applications (0 °C to -20 °C). Lower power operation for FeRAMs is obtained because of the lower programming voltages and faster programming times, as compare to Flash memory.

5. COMMERCIAL FeRAM DEVELOPMENT

Today, many major companies have announced FeRAM development programs for commercial applications, as shown in Table 1. These activities provide a critical mass of manpower and expenditures to move FeRAMs from the development stage to the commercial production phase.

Table 1. Commercial FeRAM Activities

Company	FeRAM Product
Fujitsu*	256K, 1M FeRAMs
Infineon	4M FeRAM
Matsushita*	1K FeRAM RFID Tag
NEC	1M FeRAM
Rohm*	64K FeRAM
Samsung	4M FeRAM
Sharp	1M FeRAM
STMicroelectronics	Development with IMEC
Toshiba	1M FeRAM

*production versions announced

6. CONCLUSIONS

Many advances have been made in commercial FeRAMs, including multi-level metal processing and excellent endurance and retention for both SBT and PZT. FeRAMs can replace other forms of semiconductor non-volatile memories for earth orbit and near and deep space missions. FeRAMs, when combined with radiation hardened CMOS, provide for fast programming times, low power operation, virtually unlimited endurance, and resistance to total dose, proton and high energy particle irradiation. Commercial versions of FeRAMs are being introduced into the market, but additional program funding is probably required to develop and produce radiation hardened versions.

7. REFERENCES

[1] Y. Shimada, K. Arita, E. Fujii, T. Nasu, Y. Nagano, A. Noma, Y. Izutsu, K. Nakao, K. Tanaka, T. Yamada, Y. Uemoto, K. Asari, G. Nakane, A. Inoue, T. Sumi, T. Nakakuma, S. Chaya, H. Hirano, Y. Judai, Y. Sasai and T. Otsuki, "Advanced LSI Embedded with FeRAM for Contactless IC Cards and Its Manufacturing Technology," *Integrated Ferroelectrics* (1999).

[2] S. S. Lee and N. S. Kang, FeRAM Team, Memory R&D Division, Hyundai Electronics Industries Co., Ltd. and S. C. Philpy and G. F. Derbenwick, Celis Semiconductor Corporation, "Reliability and Performance of a 64K Ferroelectric Memory" *JPL 2000 Microelectronics Reliability & Qualification (MRQ) Workshop* (2000).

High Density Packaging of Non-Volatile Memory

John C. Carson & Keith D. Gann
IRVINE SENSORS CORPORATION
http://www.irvine-sensors.com
3001 Redhill Avenue, Building 3, Costa Mesa, CA 92626-4529
Telephone: 714-444-8765   Fax: 714-444-8865
Email: jcarson@irvine-sensors.com; kgann@irvine-sensors.com

Abstract

We describe an innovation in three-dimensional (3D) packaging is heterogeneous stacks containing all of the components for a complete system or subsystem. The stack is constructed using known good die (KGD), and provides a high level of integration and interconnectivity. This packaging technology lends itself to mass storage, as well as complete computers and other complex systems. For non-volatile memory, very high densities can be achieved even through the individual chips themselves are relatively low capacity.

Introduction

The goals for the new 3D packaging technology being developed are as follows:

- Accommodate at least 50 thin (eight mil) layers.
- Start with known good dice (KGD).
- Heterogeneous stack, allowing mixed chip types.
- Allow for a high level of interconnectivity.
- A largely silicon heat conduction path out of the stack.
- Easily accommodated die shrinks.

Applications for the technology are massive memory stacks, "smart" memory modules with built-in logic, drivers, etc., and complete systems in a stack. Non-volatile memories of all types and technologies can be stacked heterogeneously. We will use FLASH and DRAM to illustrate the point.

Irvine Sensor's established die-stacking technology requires procuring the dice in wafer form, which would be logistically difficult with many die types in a stack. The other limitation of the existing technology is that it does not easily allow many different die types to be stacked together.

Neo-stacking Approach

Starting with KDG "neo-wafer" is constructed using many dice in a potting compound matrix. A standard neo-die size, just slightly larger than the largest die in the stack, is used for all dice in the stack. It is this feature that allows the stack to be heterogeneous. Blank silicon is added to open areas on layers where smaller dice are used to enhance thermal conduction between layers. The neo-wafer is metalized and thinned prior to dicing into individual neo-die. Other die types are similarly fabricated into neo-die of the same dimension. All of the necessary dice are then laminated into a single stack, with all signals to be interconnected brought out to two sides of the stack. On the top of the stack is a ceramic substrate, or cap chip, with metalization on both sides, connected through vias. Metalization is added to the two sides of the stack to complete the interconnection between dice, bringing all input/output signals to the cap chip. Figure 1 shows a cross section of a typical structure for a single die in neo form, with all signals routed to two sides. Figures 2 and 3 depict the same type of neo-die, both Flash and DRAM versions, from a top view, showing the metalization layout.

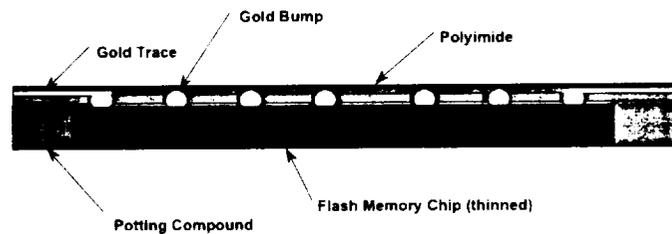**Figure 1: Typical Flash Neo-die Cross Section**
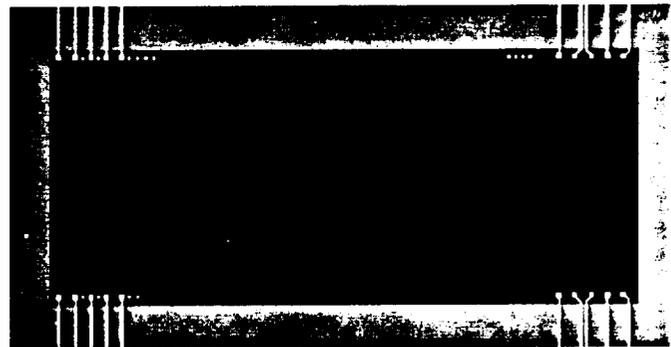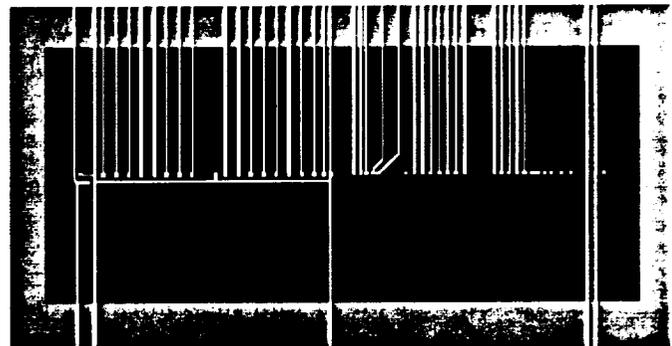

Figure 2: Flash Neo-die Layout

Figure 3: DRAM Neo-die Layout

The neo-dice shown above were fabricated in a neo-wafer, allowing for an economic batch processing methodology. The outline dimensions of the neo-wafer are the same as for standard silicon wafers. This allows standard wafer tooling and fixtures to be used for processes like metalization and thinning. Figure 4 depicts a Flash neo-wafer.

Figure 4: Flash Neo-wafer
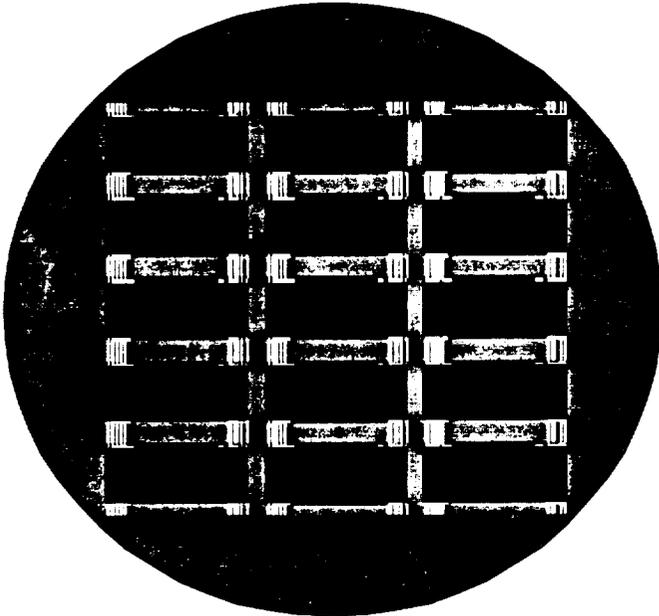
The new technology allows for multiple dice on a layer, assuming that they are small enough to fit into the standard neo-die size. A two-layer metalization structure may be necessary to interconnect the dice and exit to the busses. Figures 5 and 6 show the cross section and layout of a typical layer with multiple dice and two-layer interconnection. Note that there are two large "L" shaped pieces of blank silicon in the layer, which help transmit the heat from layers above.

Once the neo-dice are ready, and if necessary tested, they are laminated into a stack. A very thin layer of epoxy between layers holds the stack together. Being very thin (about one micron) allows heat to travel through each adhesive layer without a large temperature change. Figure 7 shows a simple four-layer neo-stack. Note that the layers are interconnected to each other and to the cap chip using bus metalization on two sides of the stack. Figure 8 shows a more complex structure with 16 layers, which uses the same construction. Despite the apparent complexity, this is still relatively simple compared with the goal of fabricating a 50-layer stack.
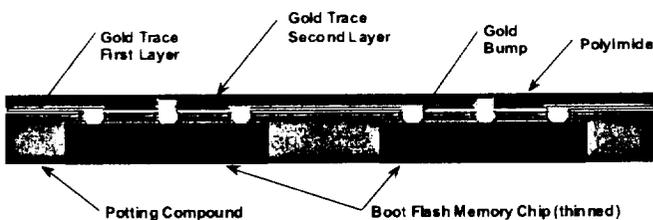
Figure 5: Boot Flash Cross Section

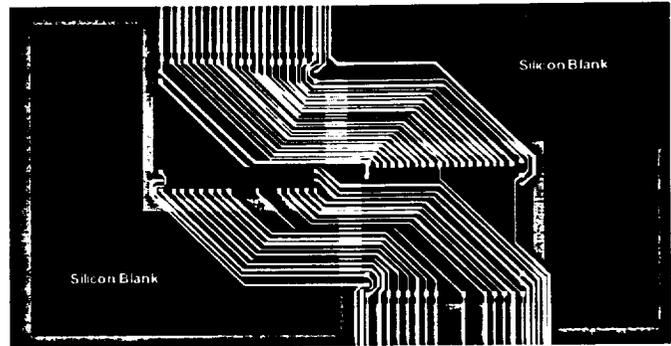**Figure 6: Boot Flash Layer Layout**

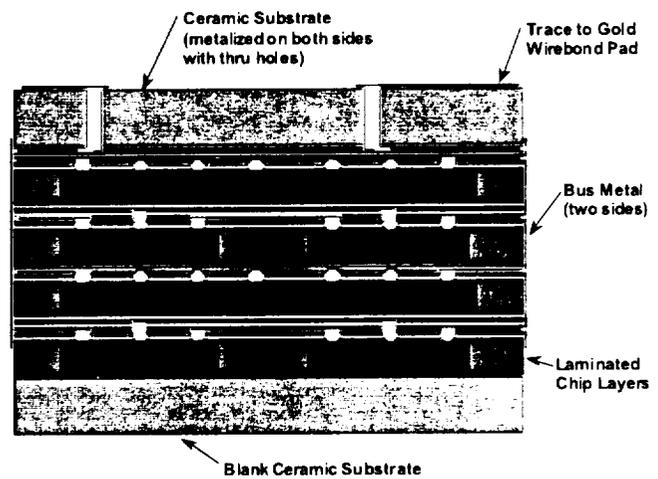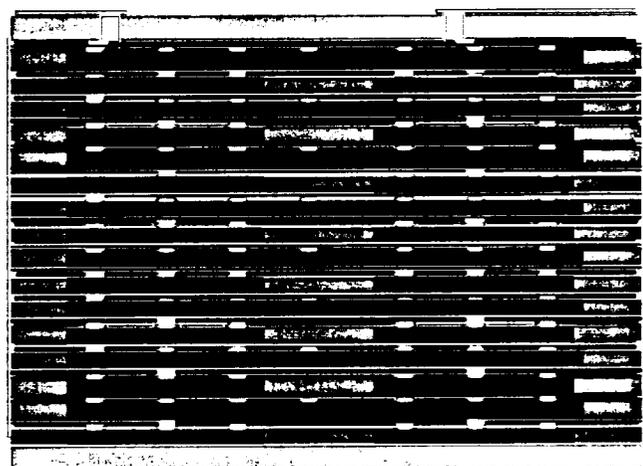**Figure 7: Simple Neo-stack Cross Section**

Figure 8: Moderate-complexity 16-Layer Stack

AHP Computer-in-a-Stack

The driving force for this development who the Advanced Humionics Platform (AHP), which was a DARPA-funded program to develop an advanced wearable computer. Irvine Sensors was teamed with The Boeing Company to develop a wearable computer with workstation capability, miniaturized to disappear into a soldier's clothing. A powerful and unobtrusive wearable computer also has many industrial and commercial applications.

Boeing is designed the total system, with Irvine Sensors developing and manufacturing the computer stacks. Each stack is a complete computer with 50 chips on 48 layers, and includes 11 chip types. All of this is in a stack that measures approximately 0.7"x 0.4" x 0.5" high. There are 32 layers of Flash memory for mass storage, eliminating the need for a hard drive. Also included in the stack are precision resisters and capacitors, fabricated on silicon and processed in the stack in the same way as a chip Table 1 shows the complexity of the AHP computer stack.

Table 1: AHP Computer Stack Layers

| NO. OF LAYERS	LAYER TYPE	ROUTING LAYERS	TOTAL CHIPS	CHIP TYPES
1	Bonding Substrate	2-sided	-	-
4	Capacitor	1	-	-
32	Flash	1	1	1
1	Flash Driver	2	4	2
1	Microprocessor	1	1	1
1	FPGA	2	1	1
1	Bus Driver	2	4	2
1	Boot Flash	2	2	1
1	IEEE 1394 Interface	2	2	2
4	DRAM	1	1	1
1	Bottom Ceramic	-	-	-

48 Total Layers   **50 Total Chips**   **11 Chip Types**

As of October 2000, Irvine Sensors has completed the fabrication of emulator stacks to validate the neo-stacking technology. The emulators are full stacks with simulated dice that have through lines. This allowed the demonstration of continuity at the cap chip interface through sequential environmental screening. Figure 9 is an isometric view of the emulator stack. The Flash memory portion was fabricated and subjected to environmental and functional testing. Irvine Sensors has begun the process of migrating the Neo-Stacking technology from low volume development to full-scale foundry service.

At 50 layers, that would allow 200 large memory chips to be in a space of 0.75"x1.2"x0.5" high. Bussing on the two long sides would allow a high level of interconnectivity. Only one or two of these large layers would be required to incorporate logic and control functions in the stack.
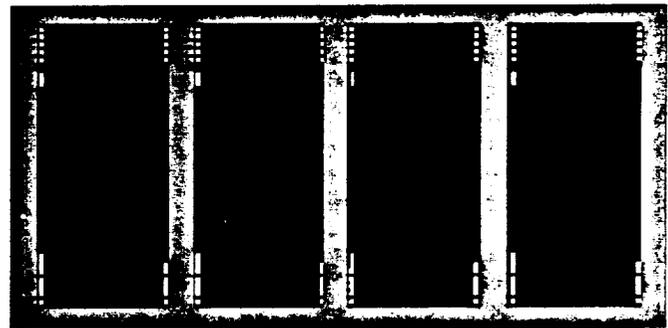
Figure 9: AHP Computer Stack
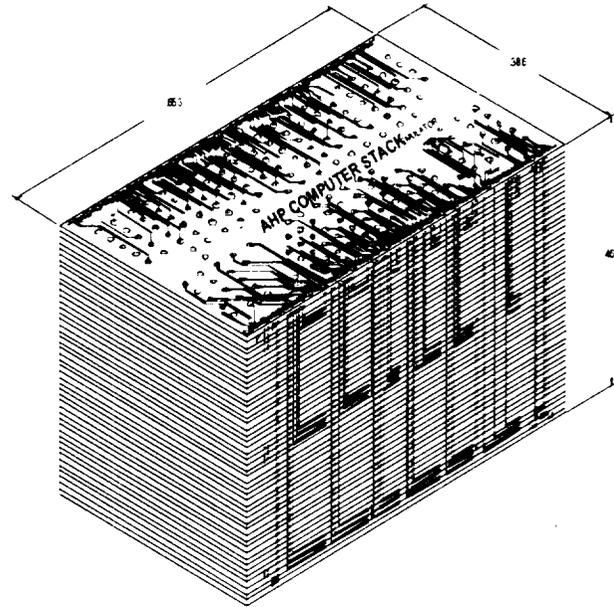
Mass Memory Applications

For mass memory applications, the design approach might be slightly different. Using the same technology, the layer footprint could be made large enough to accommodate for large non-volatile memory chips, as shown in figure 10.

Figure 10: Mass Memory Layout

Summary
Neo-stacking is a breakthrough in high density packaging technology. It allows complete systems in a cube. It also allows the combination of massive memory with extreme miniaturization and integral logic and control functions. Memory chips have been successfully demonstrated and a foundry service is being brought up.

All-Metal Magnetic RAM

E.J. Torok and R. Spitzer
IME
Minneapolis Minnesota and Berkeley Californina

1. INTRODUCTION

The factors that enter into the development of an all-metal, nonvolatile magnetic RAM, in which multilayer giant magnetoresistive films are used for all functions – storage, readout, and support electronics – are described.

2. NEED FOR NONVOLATILE MEMORIES

In the mid-1990s, the advent of portable devices – laptops, cell phones, digital cameras – made nonvolatility a huge issue. These devices all depend on nonvolatile memories to store data without the bulk, slowness, and vulnerablity of disk drives.

3. AVAILABLE NVM

The semiconductor answer to the rapidly growing applications for nonvolatile solid-state memories was flash technology. Even though hobbled by high cost, slow write time, and limited cycles, flash became the fastest-growing and most profitable memory device in the semiconductor stable.

Semiconductor memories have fundamental performance limitations that make them vulnerable to a revolutionary technology, one that would more closely meet the criteria for an ideal memory. All semiconductor memories are limited by one or more of these criteria: speed (fast read/write), low power, durability (high number of write cycles), high density, nonvolatility (maintain stored data when power is off), and low cost.

Additionally, semiconductor memories are rapidly approaching fundamental thermodynamic limits to further miniaturization. The technical difficulties facing the semiconductor industry are well summarized by Packan[1].

4. NEW MEMORY TECHNOLOGIES

A consensus is developing that the best prospect of replacing semiconductor memories, and eventually mechanical storage, with a solid-state technology in the near future lies with magnetoelectronics (ME)[2,3]. ME is a new approach to memory and storage based on magnetism, ie, on the orientation of electron spin rather than – as in semiconductors – on its electric charge.

ME is being applied to both electromechanical systems (heads for disk drives) and solid-state devices (sensors, memories, and electronics in general). The greatest economic impact of ME to date has been in the replacement of inductive heads in disk drives by heads based on giant magnetoresistance (GMR), but recently there has been a strong surge of activity in ME memories. The two main approaches to magnetic RAM being pursued are based on GMR and magnetic

tunnel junction (MTJ). Ours is a GMR technology.

Scaling of GMR technology, by contrast to semiconductor memories, is limited only by lithography. Hence the drive toward magnetic RAM. The very characteristics that are driving the most profitable and growing areas of the memory markets – namely, portability and Internet communications – are exactly what magnetic RAM is poised to deliver: nonvolatility, low power, speed, scalability, high density, and low cost. *The market forces are moving in exactly the same direction as the development of magnetoelectronics.*

GMR electronics

Initially, solid-state ME-memory development focused on hybrid ferromagnetic/semiconductor RAMs: a magnetic memory array that uses semiconductor electronics. These memories go under the generic name MRAM.

IME has developed a multifunctional, active GMR device, called the *Transpinnor* ™, that provides the foundation for all-metal electronics and enables an all-metal magnetic RAM.

SpinRAM

IME's all-metal RAM, called SpinRAM, is a nonvolatile, random-access memory in which multilayer GMR films are used for all functions: storage, readout, and support electronics. SpinRAM is a coincident-current device. Selection matrices are used to send half-select currents over a specific word line and a specific digit line to provide a full-select current at one specific bit. The digit lines serve a dual purpose in that they are also used for sensing. The sense/digit lines

are GMR multilayers with materials of different coercivities. The word lines can also be GMR films, but this is not essential.

An 8-Kbit SpinRAM chip has been fabricated. This demonstrated that the GMR electronics and the GMR memory array can be deposited and patterned concurrently. The chip is undergoing testing towards full characterization. Design of a 1-Mbit chip has been initiated.

5. CONCLUSIONS

Four significant characteristics are expected to favor all-metal over hybrid magnetic RAM.

First, silicon-technology fabrication requires a large number of masking steps, including complex ones such as ion implantation. Conversely, all-metal technology is inherently simple: fewer masking steps, no doping, scaling to lithographic limits, very little operating power.

Second, the all-metal footprint is significantly smaller than the hybrid one.

Third, an all-metal RAM is expected to be able to be miniaturized to lithographic limits; miniaturization of hybrid magnetic RAMs is likely to be limited by the semiconductor circuitry.

Finally, semiconductor processing and magnetic processing in MRAM are done separately because the former requires high temperatures, whereas magnetic fabrication is a low-temperature process. By contrast, because both GMR electronics and the memory elements are made of the same materials, the two major components are deposited and patterned concurrently on the same substrate. Compatibility issues

between CMOS and ME fabrication are avoided altogether.

6. ACKNOWLEDGEMENTS

References

1. Paul Packan, *Science*, 24 September 1999, p. 2079 "Pushing the Limits"

2. Tim Studt, *R&D Magazine*, July 1999: "Will Spintronics Replace Conventional Electronics?"

3. Mark Johnson, *IEEE Spectrum*, February 2000, p. 33: "Magnetoelectronic memories last and last..."

Advanced MRAM Concepts

James M. Daughton, Nonvolatile Electronics, Inc.

11409 Valley View Road, Eden Prairie, MN 55344 daughton@nve.com

Abstract—Two important goals of Magnetoresistive Random Access Memory (MRAM) development are to improve MRAM manufacturability and to extend MRAM density to 100 nm dimensions. One potential barrier to MRAM manufacturability is associated with the method of write selection in which two orthogonal currents in coincidence must write data, whereas each of the orthogonal currents alone cannot disturb the data. This "2D" selection method places constraints on uniformity of MRAM memory cells. Using a transistor per cell for write select greatly improves operating margins and lowers write currents. Attaining reasonable memory densities for this scheme depends on limiting the required write current in order to minimize the area of the select transistor. A second goal is to extend MRAM density to 100 nm dimensions. Use of a vertical GMR multilayer ring structure, where the data is stored in circumferentially oriented magnetizations, can extend the density of MRAM [Zhu and Prinz, Paper GB-02, 1999 Intermag]. Stability is projected for cells with inside diameters of less than 50nm. A second approach is to use Joule heating, in combination with magnetic field from a current, to write by exceeding the Neel point of an antiferromagnetic pinning layer in a pseudo-spin valve cell. Feature sizes smaller than 100 nm are projected along with decreases in required switching currents.

TABLE OF CONTENTS

1. INTRODUCTION

Development of magnetoresistive random access memory (MRAM) is progressing toward products [1,2,3,4,5]. Two potential barriers to MRAM product success are yield of the magnetic cells and the scalability of MRAM to 100 nm dimensions. Section 2 describes a new MRAM memory cell and architecture which should circumvent cell yield problems associated with traditional 2D write select schemes, which require quite tight process tolerances. In this new scheme, a select transistor per memory cell is used for writing, and a much smaller current is used for reading than for writing. This should result in substantially wider process margins, but probably at the sacrifice of density due to the size of the required transistor in the cell. This "1D magnetic select" scheme is potentially ideal for small, high performance nonvolatile RAM.

Sections 3 and 4 describe two ideas for increasing the density of MRAM apace with the industry's decreasing lithographic limits. In Section 3 an MRAM cell is described which utilizes vertical giant magnetoresistance (GMR) rather than in-plane GMR. The cell is etched into ring shapes, with the magnetizations of the GMR multilayers lying circumferencially around the ring [10]. A word line is used to provide a half-select field, while current through the ring provides the other half-select field. The advantange of the cell is improved stability by avoiding magnetic "vortices" down to very small cell sizes - at least down to inner diameters of about 300 nm. Section 4 describes a technique to increase density of MRAM by heating an antiferromagnetic pinning layer above its ordering temperature (Neel temperature). This scheme effectively deepens the energy well depth of unselected cells, and potentially will permit higher storage densities at smaller current levels.

2. 1D MAGNETIC SELECTION

Magnetic memory cells, such as magnetic cores, have traditionally used selection schemes for reading and writing which don't require a transistor or diode in the memory cell, but which do place restrictions on uniformity and "disturb" sensitivity of the cell. Figure 1 is a diagram showing a typical 2D write selection scheme. Selected cells receive both Ix and Iy currents, and are switched into the desired memory states. The currents must be selected so that Iy or Ix separately do not disturb the memory state of stored data. Bits on the same x line or y line that are not being written are subjected to "half-select" currents which tend to disturb the data. If very large currents are used to insure the writing of worst case cells, then the half-select currents are also large and tend to disturb the most disturb-sensitive cells. The half-selected memory states are also not nearly as stable as

128

stored bits, and they provide the majority of projected cell failures in time [12]. In addition to half-select currents, these cells must withstand stray fields from neighboring cells and fields from leakage currents and stray environmental fields. Thus, the requirements for uniformity and design margins present challenges in manufacturing the 2D magnetic arrays.

Most magnetoresistive memory schemes also use a 2D selection scheme for reading data. The original MRAM [6] concept and the pseudo-spin valve (PSV) [7] concept both use magnetic 2D selection schemes for reading, which introduce further disturb conditions. Magnetic tunnel junction memories (MTJ) [1,2,8] use a diode or transistor to select a memory cell for reading, and thus do not have significant disturb conditions for reading, but they still have the constraints of 2D magnetic selection for writing.
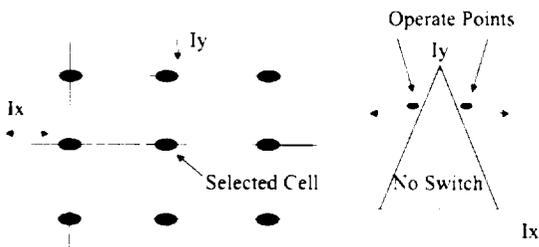
- Ix, Iy Alone Doesn't Switch Cell
- Ix, Iy Together Switch Cell

Figure 1. 2D Magnetic Selection Scheme

Figure 2 depicts a simple "1D selection" scheme for both reading and writing a magnetoresistive memory cell. A high current of either polarity (plus current for a "1" and negative current for a "0" is passed through a select transistor and through the memory cell to write. A lower current is used to generate a voltage across the cell which will be higher or lower depending on the data stored and the magnetoresistance of the cell. This voltage is then sensed and compared to a reference in order to determine the memory state.
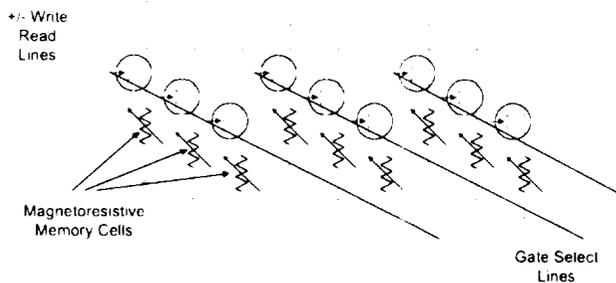

Figure 2. A 1D Magnetic Selection (Read and Write) Scheme For Magnetoresistive Cells.

Note that the transistor provides the selection of the memory cell, not the 2D magnetic switching properties of the cell. A very large current can be used to write and a small current can be used to read the cell, thus providing potentially very large margins. Of course it is important to use as small a current as will reliably write the cell so as to reduce the size of the transistor needed for selection. There are still 2D arrays of cells, but the transistors take up the burden of selection rather than placing severe constraints on the magnetic switching properties of the cell. This is why this is called a "1D magnetic selection". The scheme is quite similar to that used for DRAM where a transistor is used to write and detect charge on a capacitor.
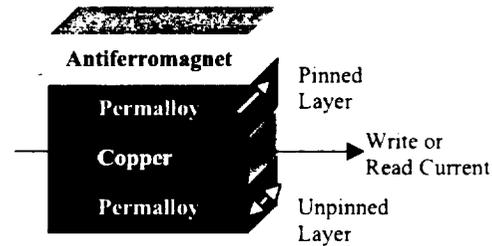

Figure 3. Cell Example For 1D Magnetic Selection.

Figure 3 shows an implementation of this concept in a cell using a "spin valve". Two magnetic films sandwich a conducting layer, and one of the two magnetic films is "pinned" with an antiferromagnet across a stripe etched from the materials. A magnetic field created by a current through the stripe can be used to magnetize the unpinned magnetic film in either of two directions, depending on the direction of the current. Then a smaller current through the stripe can be used to sense the value of resistance - higher if the films are oppositely magnetized and lower if they are magnetized in the same direction. A large current can be used for writing without disturbing other cells, and a much lower current can be used for sensing. This would suggest large margins. In practice, this cell would have difficulties with demagnetization fields in cells with micron dimensions and smaller.
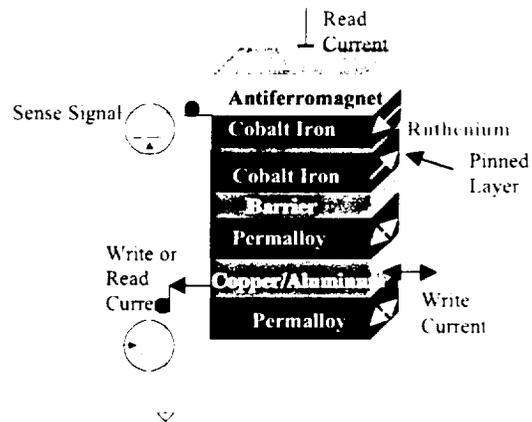

Figure 4. 1D Magnetic Write Selection Cell With Tunneling Readout.

Figure 4 shows a variation on the concept that uses spin dependent tunneling for a higher signal output and minimizing demagnetizing effects. A sandwich stripe with neither magnetic film "pinned" can be magnetized with a current through the sandwich into either of two magnetic states depending on the direction of current. With no magnetic field applied, the magnetizaions of the films are antiparallel and lie across the axis of the stripe. A tunnel barrier is deposited on top of the sandwich, and a pinned synthetic antiferromagnet is deposited on top of the barrier. The top layer of the synthetic antiferromagnet is pinned with an antiferromagnet. The synthetic antiferromagnet is comprised of two magnetic layers sandwiching a thin ruthenium layer, a structure which has very strong antiparallel coupling. The synthetic antiferromagnet thus produces very little stray magnetic field, and similarly, the unpinned sandwich also produces very little stray magnetic field. The spin tunneling current between the top layer of the sandwich and the pinned layer is then either higher or lower depending on the direction of magnetization of the top layer. Spin dependent tunneling has demonstrated about 40% magnetoresistance with 100 mV across the barrier, and thus a signal of approximately 40 mV could be observed. The barrier resistance is usually large enough to limit the read current to about 20 µA, while the current needed to set the magnetization in the cell (write) should be about 2.5 mA for a cell on the order of a micron width. For those values, the disturb current would be about 1/100 of the write current.

Table 1 gives a qualitative comparison of memory properties for spin dependent tunneling and pseudo-spin valve memory cells (based on limited published data) with the potential properties of memories based on this 1D cell. The major potential advantages for these memories are much lower write currents (power) and manufacturabililty. Memory performance should be as high as for MTJ memories, but because of the size requirement for the write transistor, densities will be lower, especially when compared with the pseudo-spin valve based memories. However, for many small, niche memories, the overhead space required for pads and peripherial electronics is much larger than the space required by the memory cells. In such applications the 1D selection scheme could be especially useful.

	PSV	MTJ	1D Cell
Write Current	25 mA	25 mA	2.5 mA
Ease of Mfg.	Harder	Harder	Easier
Speed	High	Medium	High
Density	High	Medium	Lower

Table 1. Comparison of Potential Properties of 1D Cells with some projected properties of PSV and MTJ cells.

3. VERTICAL RING GMR CELLS

One of the problems encountered in high density MRAM cells has been magnetic anomalies or vortices [9]. These vortices cause inconsistant magnetic behavior, and may result in the inability to write or read a cell correctly. Magnetic simulations and measurements have confirmed this behavior. A structure which may circumvent this problem [10, 11] is shown in Figure 5. Magnetic films of a few nanometers thickness are separated by thin copper layers, and alternating magnetic layers are of two significantly different thicknesses. When a bit current is passed through the vertical stack, a circumferential magnetic field is created that can tend to magnetize these magnetic films, with the thinner magnetic layers switching at lower currents (fields) and the thicker layers switching at higher currents (fields). With the layers magnetized in one sense, the resistance of the stack is lower than when the layers are oppositely magnetized. The vertical GMR is in general larger than in-plane GMR, so that for a given current through the stack, the signal is quite usable, even though the cell itself generally has relatively low resistance due to its shape.
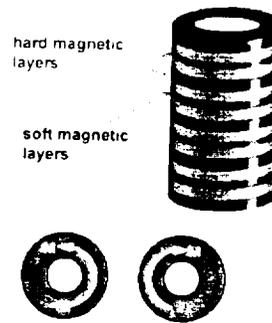

Figure 5. A Vertical GMR Cell and Circumferential Remnant Magnetizations.

In order to attain a 2D read/write select organization, a pair of parallel word lines lying on the outside edge of the cell carry currents in opposite directions, and an orthogonal pair of lines carrying similar currents is placed on the opposite side of the cell, as shown in Figure 6. With currents flowing through both pairs of word lines, a field configuration shown in Figure 7 is created. This field is a "tipping" field which lowers the bit current thresholds for switching. Thus, a 2D selection scheme is possible where the word currents by themselves or the bit currents by themselves cannot write the hard (thick) film layer, but the combination of the two currents can.
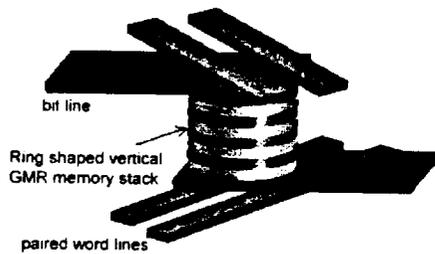
Figure 6. Vertical GMR Memory Cell Showing Bit Lines (Bit Current Direction Determining Data Written) and Word (Tipping Field) Lines.

One way read out data from such a cell is to switch the soft (thin) layer from one state to the other while observing the resistance. At some intermediate value of word current and sense current where the soft (thin) layer can be switched, but the hard (thick layer is not), the resistance of the sense line is observed as the sense current is reversed. In this way, the sense of the hard layer can be determined. In Figure 8, this would correspond to operating in the flat top region where the soft film is switched back and forth. The resistance of the cell is higher when the magnetization of the soft layer is antiparallel to the magnetization of the hard layer, and lower with a parallel orientation.
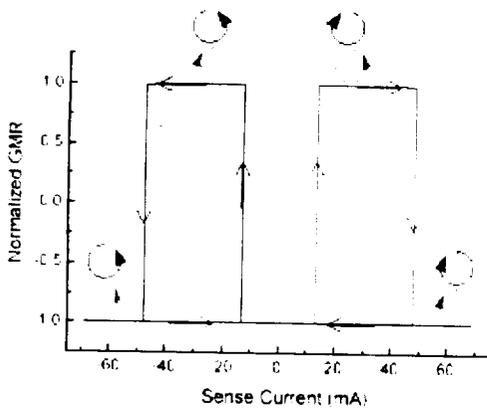

Figure 8. Magnetoresistance as a Function of Sense Current.

The primary advantage of this cell is the potential density which it can achieve. The storage mode, according to extensive numerical simulations, does not create vortices for cell sizes down to less than 100 nm diameter (approximately 30nm inside diameter), which is smaller than demonstrated PSV and MTJ cells. Data confirming this small achievable size for vertical GMR cells has not been published.

4. NEEL POINT WRITTEN CELLS

Another challenge for very high density MRAM is cell stability at nm dimensions. As the cell size shrinks and

the volume, V, of magnetic materials gets smaller, thermal agitation can cause a cell to lose data. (This same problem gives rise to the so called "superparamagnetic limit" spoken of in recording technology". If Ht is the switching threshold of a half-selected cell and Ms is the magnetic moment, then the depth of the energy well associated with that switching threshold in the half-select state is proportional to Ms*Ht*V. For low error rates the ratio of this energy well depth to kT should be at least 55 [12]. Assuming a maximum operating temperature of 350 K and a cell size of 100nm X 100nm X 2 (nm), one finds that Ht must be about 166 Oe. and the threshold of a non-half-selected cell must be several times that. When enough current is applied to get this magnitude of field in a nm sized cell, thermal heating is a problem, and kT becomes even larger, and the higher densities are even harder to achieve.

An approach to improving the density of MRAM is to make use of heating effects in combination with magnetic fields from currents to switch cells which use antiferromagnetic pinning of a ferromagnetic layer. Figure 9 illustrates a thin ferromagnetic film which is deposited in contact with an antiferromagnet. After pinning, very large magnetic fields (several thousands of Oe) cannot permanently reverse the pinned direction if the temperature is significantly below the Neel (ordering) temperature of the antiferromagnet. This property could be used in many memory cells to obtain a deep energy well for stored data, and provided heat can be applied to the cell for writing, the writing currents may not have to be very large.
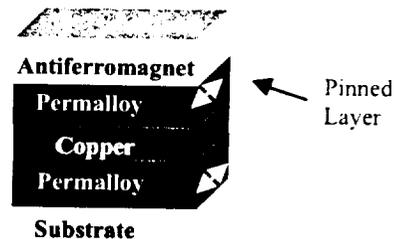

Figure 9. A Permalloy Layer Pinned With An Aniferromagnet. A second permalloy layer and a copper layer can be used to complete a PSV structure.

If the pinned layer in Figure 9 can be substituted for the storage layer in a PSV [7], and if the heating effects of the word and digit currents can heat the ferromagnet above the Neel temperature of the antiferromagnet, then the cell can be written in exactly the same way as a normal PSV, but the cell would have a decidedly higher stability.

In the PSV cell shown in Figure 10, current through the sense line acts as a tipping field and current in the word line in combination with that tipping field reverses the storage layer if the word current is large enough, but reverses only the soft layer at a smaller word current.
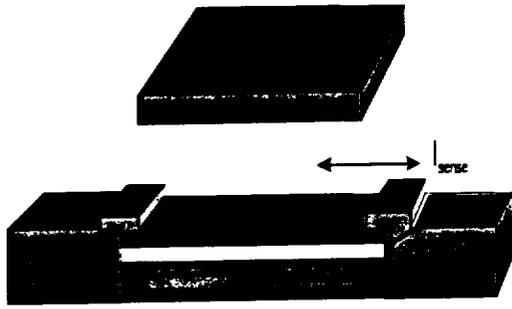
Figure 10. A Psuedo Spin Valve (PSV) Cell.

Pinned ferromagnetic films were fabricated at NVE as storage layers in pseudo-spin valve cells in order to obtain experimental verification of switching using a combination of heating and magnetic field.. The pinned layer was cobalt-permalloy pinned with an iron manganese antiferromagnet, and the unpinned read layer was cobalt-permalloy. Repeatable switching was observed. A thermal time constant of about 3 ns was demonstrated for a silicon nitride dielectric thicknesses of 35 nm.

5. CONCLUSIONS

One approach to making a producible. high performance niche memory and two approaches for extending the density of MRAM to nm dimensions were described. It should be noted that these techniques could be used in combination. There are undoubtedly many more possibilities for improving MRAM density. performance, and producibility that will come to light in the next few years.

6. ACKNOWLEDGEMENTS

7. REFERENCES

[1] S. Tehrani. B. Engle. E. Chen. M. DeHerrea. M. Durlam, and J. Slaughter, " Recent Developments in Magnetic Tunnel Junction MRAM". Paper DA-02 presented at Intermag 2000 in Toronto, May, 2000.

[2] R. Scheuerlein. W. Gallagher, et al. "High Performance Demonstration of Magnetic Tunnel Junction Random Access Memory". Paper FB-01 presented at the MMM Conference. November, 1999.

[3] "USTC Puts Magnetoresistive RAM to Commercial Use". Electronic Buyers News, July 24, 2000.

[4] "MRAM: New Ways to Store More". Hewlett-Packard Web Site.

[5] "Motorola Demonstrates Magnetoresistive RAM". EE Times Monday, May 15, 2000.

[6] J.Daughton, "Magnetoresistive Memory Technology". Thin Solid Films 216. p 162 (1992).

[7] A. Pohm. J. Anderson, R. Beech. and J. Daughton, "Bias Field and End Effects on the Switching Thresholds of Pseudo Spin Valve Memory Cells". IEE Trans. Magn. 34. No. 4. p 3280 (1997).

[8] J. Daughton. "Magnetic Tunneling Applied To Memory". J. Appl. Phys. 81. No. 8. p 3758 (1997).

[9] J. Gadbois and J. Zhu. "The Effect of End and Edge Shape on the Performance of Pseudo-Spin Valve Memories", IEEE Trans. Magn. 34, No. 4. p 1066 (1008).

[10] D. Bussman. G. Prinz. S.Cheng. J. Zhu. Y. Zheng. J. Daughton. R. Beech. D. Wang, and R. Womack. "Current Perpendicular-To-Plane GMR For Magnetoresistive RAM, Paper GA-03 presented at the 1999 Intermag Conference in Kyongiu. Korea. May. 1999.

[11] J. Zhu. Y. Zheng, and G. Prinz. "Ultrahigh Density Vertical Magnetoresistive Random Access Memory". J. Appl. Phys. 87. N. 9. p. 6668 (2000)

[12] R. Beech. J. Anderson. A. Pohm. and J. Daughton, "Curie Point Written Magnetoresistive Memory". J. Appl. Phys. 87, N. 9. p 6403. May, 2000.

REPORT DOCUMENTATION PAGE

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE November, 2000	3. REPORT TYPE AND DATES COVERED JPL Publication

4. TITLE AND SUBTITLE
Proceedings
Non-Volatile Memory Technology Symposium 2000

5. FUNDING NUMBERS
C-NAS7-1407

6. AUTHOR(S)
Nazeeh Aranki (editor)

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Jet Propulsion Laboratory
California Institute of Technology
4800 Oak Grove Drive
Pasadena, CA 91109-8099

8. PERFORMING ORGANIZATION REPORT NUMBER

JPL Publication 00-15

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
National Aeronautics and Space Administration
Washington, DC 20546-001

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT

Subject Category: 33   Distribution: Nonstandard
Availability: NASA CASI (301) 621-0390

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)

This publication contains the proceedings for the Non-Volatile Memory Technology Symposium 2000 that was held on November 15-16, 2000 in Arlington, Virginia. The proceedings contains a wide range of papers that cover the presentations of myriad advances in the nonvolatile memory technology during the recent past including memory cell design, simulations, radiation environment, and emerging memory technologies. The papers presented in the proceedings address the design challenges and applications and deals with newer, emerging memory technologies as well as related issues of radiation environment and die packaging.

14. SUBJECT TERMS

Non-Volatile Memory   Embedded Flash Memory
Metal Magnetic RAM   Flash Memory Cell
Radiation Effects/Issues FeRAM Technology

15. NUMBER OF PAGES
135

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassfied	Unlimited