<***Invited Talk*** *for 15th International Conference on Ion Implantation Technology, 2004*>

Advanced Non-Volatile Memory Devices with Nano-Technology

Chih-Yuan Lu and Chih-Chieh Yeh

Macronix International Co.,Ltd, No.16, Li-Hsin Road, Science Park, Hsin-Chu, Taiwan, ROC

Email:cylu@mxic.com.tw, Fax:886-3-5785336, Phone:886-3-578-6688 ext:76667

Abstract

With increasing challenges of scaling conventional floating gate flash memory technology, alternative non-volatile memory technologies (Table 1) are under investigating. In this paper, the scaling limitations of the conventional floating gate technologist are reviewed first. In the following, nitride based flash memory cells are introduced, which is a breakthrough flash memory device architecture solution. In the third part, new demands and applications of flash memory technologies are discussed. Finally, several new non-volatile memories are summarized, which provide new opportunities for further scaling and cost reduction in nano-scale era.

Table 1 Summary of emerging non-volatile memories.

Transistor Vt shifts	Charge Displacements	Resistance Change
1. Floating gate 2. Nitride trap 3. Nano-crystal	1. Crystalline Ferroelectric 2. Polymer Ferroelectric	1. Magnetic: GMR or MJT 2. Phase Change 3. Polymer ionic transport

Introduction

Today, flash memories find wide applications and are considered as a technology driver for semiconductor industry in the next decade. It can be classified into two major markets: code storage application and data storage application (Fig.1). NOR type flash memory [1] is most suitable for code storage application, such as cellular phone, PC bios, and DVD player. NAND type flash memory [2] has been targeted at data storage market, which is an emerging application such as PDA, memory cards, multi-media audio, and digital still camera. Fig.2 discloses the memory market and flash memory share increased rapidly in the last few years. Fig.3 shows the NOR/NAND annual revenue. Although market size of data storage is still smaller than that of code storage market before 2004, the growth rate points out that flash memories for the data storage will become a major semiconductor product in the near future.

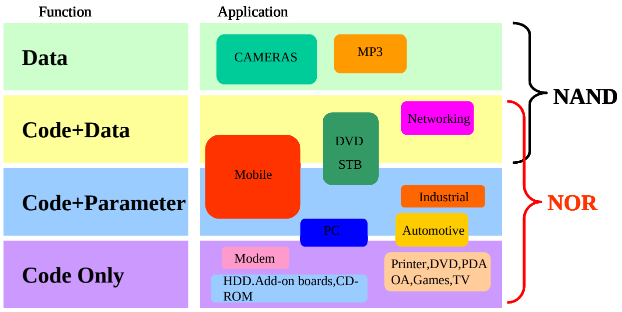


Fig.1 Major application of flash memories (Web Feet Inc. 2003).

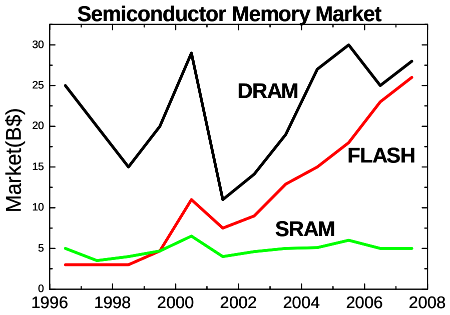


Fig.2 Forecast of memory market share (Web Feet Inc. 2003).

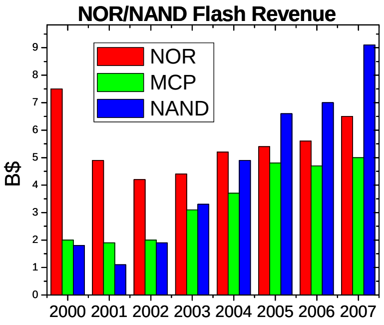


Fig.3 Forecast of NOR, NAND and MCP revenue (Web Feet Inc. 2003).

Floating gate flash scaling considerations

NOR and NAND share the same floating gate flash memory structure as shown in Fig.4. Table 2 shows the ITRS roadmap [3] for flash memory cells. NOR has good visibility into 90nm and 65nm generation. Current projection shows that scaling continues at 45nm node but is challenged to meet the goal of 50% cell size shrinkage [1,4]. NOR has two scaling limiters. One is on the non-scalability of tunnel oxide and inter-poly ONO due to reliability concerns. The other is caused by the channel hot electron programming, which requires internal voltage to more than 8 volts, and imposes the limit of the cell gate length [4]. Besides, the process complexity increases dramatically to shrink the memory cell size as shown in Table 3, which makes the cost ineffectiveness.

Fig.5 shows the NAND technology cell size trend. NAND meets another scaling limitation caused by floating gate itself. Owing to the stray capacitance coupling, gate coupling ratio will drop dramatically when the wordline spacing is smaller than 40nm even under an ideal scenario of scalable floating gate thickness [5]. Besides, floating gate interference is another scaling limit of NAND. Vt shift is caused due to the Vt change of the adjacent cells. It results from capacitive coupling via parasitic capacitors around the floating gate, which become prominent as device size/spacing is scaled down. It is very likely that the practical limit of NAND flash scaling is at 30nm technology node.

Instead of process scaling, one important new innovation in cost reduction is multi level charge storage (MLC). So far, products with 2 bits per cell have been shipping for years. Looking into future, it is possible to have 16 levels storage giving 4 bits per cell, in which case some level of error correction will be required [6]. Although MLC can double the memory density, scaling limitations of conventional floating gate memory cell mentioned above are coming in the near future. New memory concepts are urgent to maintain Moore's low cost learning curve.

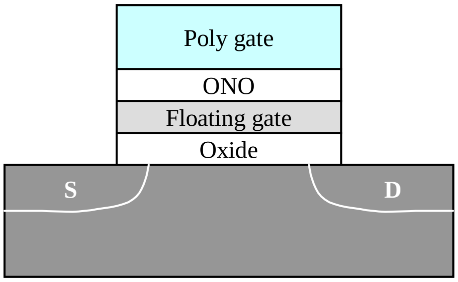


Fig.4 Cell structure of floating gate flash memory.

Table 2 NOR and NAND technology roadmap in ITRS 2002.

Year		2001	2004	2007	2010	2013
Technology node (nm)		130	90	65	50	35
NOR	Cell Size (F^2)	10	10	11 ~14	12 ~15	13 ~16
	Cell Size (μm^2)	0.16	0.081	0.053	0.034	0.018
NAND	Cell Size (F^2)	5.5	5.5	4.5	4.5	4.5
	Cell Size (μm^2)	0.093	0.045	0.019	0.011	0.006

Table 3 Demonstrated NOR device/process features [4].

Process\Tech Node	**1μm**	**0.8μm**	**0.6μm**	**0.4μm**	**0.25μm**	**0.18μm**	**0.13μm**
Tunnl ox (A)	120-110	120-110	110-100	110-100	100-90	100-90	100-90
Interpoly dielectric (A)	300-250	250-200	200-180	180-160	160-140	160-140	160-140
Weff/Leff (μm)	0.6/0.4	0.5/0.35	0.45/0.32	0.35/0.25	0.29/0.2	0.19/0.14	0.13/0.1
GCR	0.6	0.55	0.55	0.6	0.6	0.6	0.6
Iread (uA)	100-75	75	60	50-60	50-60	50-60	50-60
Self-align source	NO	NO	Yes	Yes	Yes	Yes	Yes
STI/SA-STI	NO	NO	NO	NO	Yes	Yes	Yes
Dual/STI	NO	NO	NO	NO	NO	NO	Yes
Unlanded/SA Contact	NO	NO	NO	NO	NO	NO	Yes
Poly CMP	NO	NO	NO	NO	NO	NO	Yes
Local Interconnect	NO	NO	NO	NO	NO	NO	Yes
Active ratio=WL/cell size	0.2	0.21	0.23	0.28	0.29	0.28	0.17

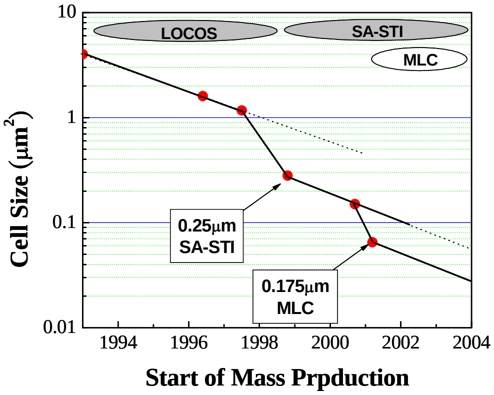


Fig.5 NAND technology trend [2]. Cell size is scaled down to 60% by using SA-STI, and to 50% by using MLC.

Nitride based trapping storage flash memory

SONOS flash memory cell (Fowler-Nordheim tunnel program by electron and direct tunnel erase by hole [7]) has been proposed for years. As shown in Fig.6, the carriers are stored in the traps of the nitride layer between top and bottom oxides. SONOS cell offers several advantages over conventional floating gate memory cell: simple process, ease of manufacturing, no erratic bit, not sensitive to oxide defects, and no floating gate coupling effect. However, the cell retention is still an issue until now, and its large cell size ($6F^2$) and slow program/erase speed limit its applications. Recently, SONOS cell has evolved into an 2 bits storage architecture by utilizing the localized charge trapping effect of nitride, which enables a memory cell to hold twice as much data as standard memory cell, without compromising device endurance, performance or reliability. Lot of vendors including AMD, Fujitsu, Infeneon, Saifun, and Macronix develop their next generation flash technologies based on 2-bit SONOS cell structure. Among numerously proposed cell architectures, NROM [8] and PHINES [9] are the most promising technologies.

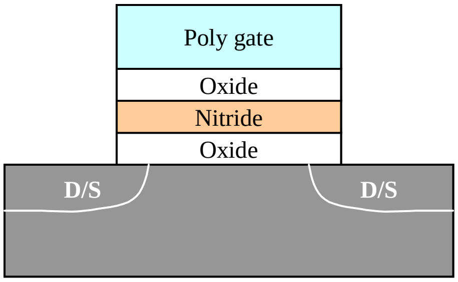


Fig.6 Cell structure of nitride storage flash memory.

NROM flash cell is to utilize SONOS cell structure (Fig.6) for physical 2 bits per cell storage. NROM programs its memory cell by channel hot electron injection as conventional NOR-type floating gate memory does, which is suitable for code storage applications. NROM holds 2 physical bits per cell above the source and drain junction, which is more reliable alternative to MLC solutions in the floating gate memory cell. MLC products suffer from performance and reliability concerns inherent when detecting between multiple charge levels.

PHINES flash cell (proposed by Macronix) also uses SONOS cell structure. Single cell operation with $5F^2$ per cell and $2.5F^2$ per bit is proposed [9]. PHINES uses band-to-band hot hole for 2-bits-per-cell programming and FN injection for erasing, which realizes ultra low power operations. Accordingly, its small bit size, simple process, high cell performance, and low power operation can meet the requirements of data storage applications.

NROM demonstrates its high performance for code application while PHINES shows its superior advantages in data storage application. As the conventional floating gate flash memory faces the scaling limitations in the near future, nitride based trapping storage memory cell may be an evolutionary concept with shorter learning curve and faster time to market.

MCP, integrated code and data, and embedded flash

Advanced cell phone, PDA and multi medias require higher memory density and NOR is insufficient to meet the demand. Although NAND has cost effectiveness, it is generally not possible to “boot” from NAND without using a dedicated state machine or controller because of its indirect interface. “Combo memory” is essential for these applications. However, technology integration of NOR and NAND is quite difficult due to incompatible process and array architecture in the conventional floating gate structure. MCP (multi-chip packaging) technology [10] is developed for this issue (Fig.7). MCP technology can house multi chips, and multi functions in a single package. Vendors can combine SDRAM, SRAM, NOR flash memory, NAND flash memory, controller chip or other functions to serve their requirements. MCP is attractive in “systems solution” since the elements are proven, design-and-debug is fast, get to market faster, and the cost is effective in smaller volumes. However, for “combo memory” applications, mass volume, bad die issue, and testing cost are major challenges.

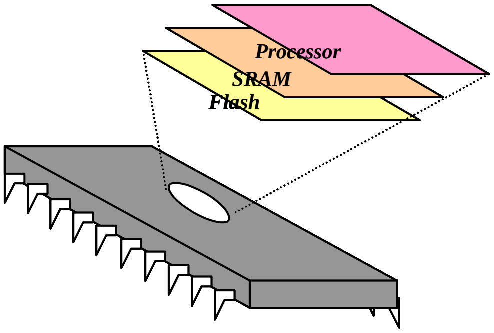
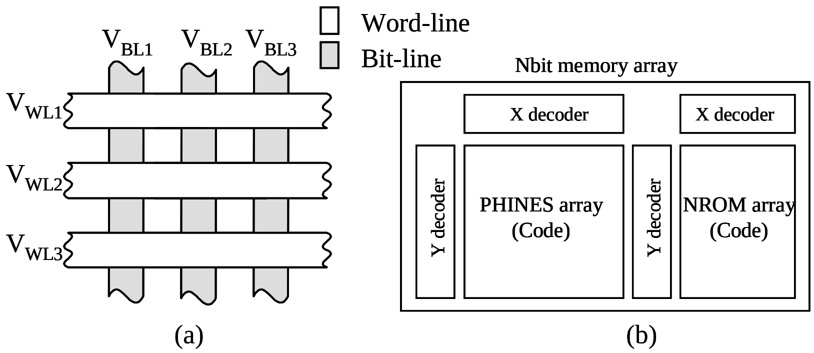


Fig.7 Illustration of MCP.

SONOS technology provides another niche for “combo memory” application since it can integrate code and data flash memory cell in one fabrication process. NROM for code application and PHINES for data application can be integrated together easily since they share the same cell structure and array architecture as shown in Fig.8 (a) [11]. The cell performance, electrical specifications and memory density of NROM and PHINES can be optimized separately to meet various product demands (Fig.8 (b)). Accordingly, SONOS technology can either encroach traditional NOR and NAND flash market or exploits new market.

Embedded flash memory (flash memory + logic) is another emerging technology for wireless and portable products. Compare to package level integration (MCP or SIP (system in package)) mentioned above, embedded flash technology provides a silicon level integration (SOC, system on chip)) solution, which is capable of combining advanced logic and non-volatile memory. For many, the higher costs associated to SOC or embedded memory are the main problem. Mask costs are an issue. Each increases cost and reduces yield. Another important demand of embedded applications is low voltage operation. Conventional nonvolatile embedded memory needs high peripheral voltages to operate memory module. High voltage pumper and driver will



(a)

(b)

Fig.8 (a) Layout representation of Nbit cells arranged in a virtual-ground array.
(b) Schematic representation of the combo (NROM and PHINES) memory array architecture.

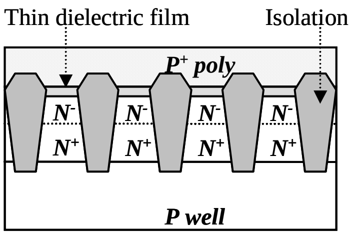


Fig.9 Schematic representation of PREM cell structure.

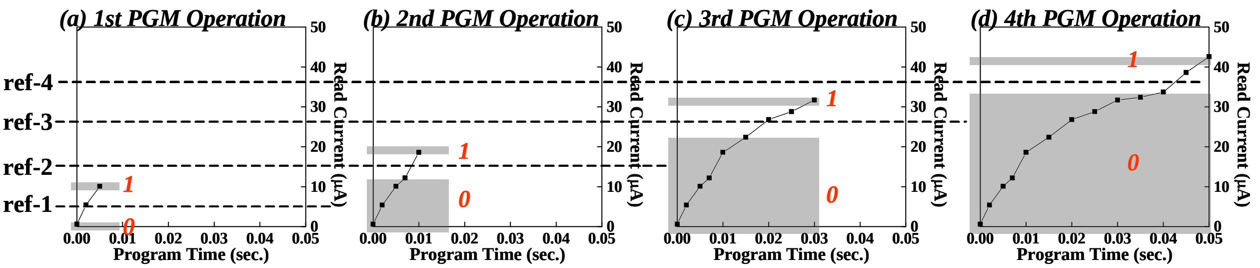


Fig.10 An exemplary illustration of a PREM cell to realize MTP. (a), (b), (c) and (d) represent the schematic representations of 1st, 2nd, 3rd and 4th time program (PGM) operations and ref-1, ref-2, ref-3 and ref-4 are the reference currents for each time program operation.

consume chip area and complicate the fabrication process. Macronix developed a novel non-volatile memory cell named “PREM” as shown in Fig.9 (Programmable Resistor with Erase-less Memory [12]). PREM uses a new “erase-less” algorithm (Fig.10) and the progressive breakdown in ultra-thin oxide. No extra mask is needed with CMOS standard process, which is specifically suitable for SOC applications. MTP (multi-time programming), MLC (multi- level cell), non-volatility, and low voltage operation (<±3V) are realized. PREM provides low cost and its high performance can meet most demands of embedded applications.

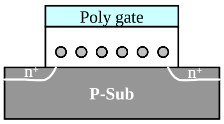
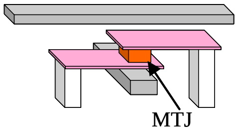
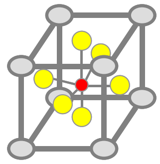
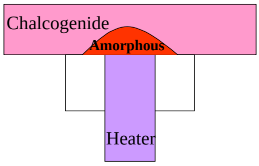
Novel revolutionary non-volatile memory cells

Table 4 summarizes the emerging non-volatile memories with revolutionary materials. Table 4 compares their cell characteristics. Among them, nanocrystal memory [13] based on discrete storage nodes has the potential for pushing further the scaling limits of conventional floating gate memory. Nanocrystal memory has the advantages of (1) higher scalability with scalable tunnel oxide thickness; (2) fully CMOS compatible process for embedded application; (3) not sensitive to oxide defects with good intrinsic performance. However, some issues are needed to be overcome: (1) low threshold voltage shift due to limited gate surface coverage; (2) fluctuation of electrical characteristics due to the dot density from one device to the other; (3) metal contamination if metal nanocrystal is used instead of silicon nanocrystal; (4) bad retention after cycling.

Magnetoresistive Random Access Memory (MRAM [14]) has been developed for years. The memory cell is composed by 1 MJT with 1 transistor. MRAM programs via switching the magnetic polarity by electric current and senses the resistance change. Its low voltage operation, high program speed, fast read, and unlimited cycles attract memory vendors. However, the writing current is too large (~8mA) to realize high density MRAM. Larger cell size, scalability, and power are the limiting factors of MRAM toward next generation flash memory cell.

FeRAM (ferroelectric random access memory [15])

Table4 Emerging non-volatile memories for next generation. Including nano-crystal memory cell, MRAM cell, FeRAM cell, and Phase change memory cell [16].

Cell	Nano-crystal	MRAM	FeRAM	PCM
Structure				
Endurance	10^6	$>10^{13}$	$>10^{13}$	$>10^{13}$
Cell write time (ns)	1000	50	30	30
Cell write current (μA)	10^{-3}	8000	5	<2000
Cell Size (F^2)	4?	20	40	6
Main advantage	Low cost Low power	Fast read Fast write Unlimited cycles Non-destructive read	Fast read Fast wite Unlimited cycles	Fast read Fast wite Unlimited cycles Non-destructive read
Main issues	Small Vt window Fluctuations of electrical parameters Retention	Write current Large cell size	Destructive read Retention Material texture	Write current Large cell size

has many desirable features, including low voltage operation, fast program/erase speed and nonvolatility. However, so far no one has realized truly FeRAM technology since FeRAM exhibits a retention problem (the best retention time has not exceed several days). Leakage current and depolarization field are the two major causes of the bad retentivity. Despite numerous researches and attempts, these two issues are very difficult to eliminate, which limits its applications.

Chalcoganide phase change memory has been studied as a candidate for next generation flash memory [17]. Electric current is passed form a heater to the chalcoganide and local joule heating is used to change the phase around the contact region. Fast write, fast read, and high endurance are its features while the time required for switching is typically less than 30ns. Higher current (>1mA per cell) for operation is its issue, which needs larger contact area and larger driving transistor (X-decoder) to handle the current. Higher circuit overhead and larger die area are challenges, which makes the cost of phase change memory ineffective.

Conclusion

Flash memory has come of age as a mainstream memory product, and its technologies and markets will become more diversified. As the conventional floating gate memory faces the scaling limitations in the near future, lots of memory technologies has been studied as a candidate. Except for nitride based trapping memory, most emerging memories are not commercially available for fabricating yet. Among nitride based trapping memory cells, NROM and PHINES provide superior performance and can be integrated together for various demands, which have the potential to dominate in future generations. Otherwise, our newly invented PREM cell can meet the requirements of low voltage, non-volatility, and simple process for SOC, embedded, and high-density storage applications.

Reference

[1] S. Lai, Future Trends of Nonvolatile Memory Technology, December 2001.
[2] S. Aritome, *IEEE IEDM Tech*. Dig., 2000, p.763
[3] Process Integration, Device and Structures and Emerging Research Devices, ITRS, 2002 edition.
[4] Sam Pan et al, *SSDM*, 2002, p.152
[5] J.D. Lee et al, *IEEE EDL*, 2002, p.264
[6] S. Lai, *IEEE IEDM Tech*. Dig., 1998, p.971
[7] I. Fujiwara et al, *IEEE IEDM Tech*. Dig., 1998, p.995
[8] B. Eitan et al, *IEEE EDL*, 2000, p.543
[9] C.C. Yeh et al, *IEEE IEDM Tech*. Dig., 2002, p.931
[10] Levinstein H., *ISSCC*, p.224, 1987.
[11] C.C. Yeh et al, to be published
[12] C.C. Yeh et al, to be published
[13] R. Muralidhar et al, *IEEE IEDM Tech*. Dig., 2003, p.26.2.1
[14] Tehrano S. et al, *IEEE Trans. on Magnetics*, 1999, p.2814
[15] Hiroshi I., *IEEE IEDM Tech*. Dig., 2003, p.10.3.1
[16] Herman E. Maes et al, *NVSM workshop*, 2003, p.2
[17] S. Lai, *IEEE IEDM Tech*. Dig., 2003, p.10.1.1