Available online at www.sciencedirect.com
SCIENCE DIRECT®
Materials Science in Semiconductor Processing 7 (2004) 349–355
MATERIALS SCIENCE IN SEMICONDUCTOR PROCESSING

Non-volatile memory technologies: emerging concepts and new materials

Roberto Bez*, Agostino Pirovano

STMicroelectronics, Non-Volatile Memory Technology Development, Central R&D, Via C.Olivetti 2, 20041 Agrate Brianza (Milan), Italy

Abstract

The current mainstream, based on the Flash technology, is expected to be the reference technology also for the next few years. Nevertheless Flash has technological and physical constraints that make more difficult their further scaling. In this contest there is the industrial interest for alternative technologies that exploit new materials and concepts to go beyond the Flash technology, to allow better scaling, and to enlarge the memory performance. The main emerging non-volatile memory technologies based on inorganic material, like ferroelectric memory (FeRAM), magnetoresistive memory (MRAM) or phase change memory (PCM) and the main innovative concepts based on organic material like ferroelectric or conductance switching polymer will therefore be analysed. We will then focus the attention on PCM technology as one of the best candidate as next-decade non-volatile memory technology, covering the main characteristics and presenting the latest development results.

1. Introduction

In the last decade the impressive growth of the portable systems market (palmtop, mobile PC, mp3 audio player, digital camera and so on) has attracted the interest of the semiconductor industry on non-volatile memory (NVM) technologies for both code and mass storage applications. Since the demand for mobility applications is the main driver behind NVM technologies and components, ever higher capacity and system performance, lower power consumption, smaller form factor and lower system costs are thus required. Floating-gate Flash memories can satisfy these needs in the short run. In the long run, several of the alternative NVMs have the potential to do so.

The current NVM mainstream is based on the floating gate (FG) Flash technology, which has a nearly 100% market share, and it is expected that NOR and NAND Flash will be the high volume NVM production for the rest of this decade. Although the Flash memory technology has been able to follow the evolution of the semiconductor roadmap [1,2] for more than 10 years and it is expected to scale with the same trend for the next technology nodes, there are physical limitations to be faced and the downscaling beyond the 45 nm technology node is still considered critical.

In this contest there is the industrial interest for alternative technologies that exploit new materials and concepts to go beyond the Flash technology, to allow better scaling, and to enlarge the memory performance. A total of more than 30 NVM technologies and technology variations compete for a piece of the fast growing NVM market, many of them aiming to replace DRAMs. However, long-term survival will be ensured only by those who will succeed in transforming

*Corresponding author. Tel.: +39 039 603 5572; fax: +39 039 603 5233.
E-mail address: roberto.bez@st.com (R. Bez).

1369-8001/$ - see front matter © 2004 Elsevier Ltd. All rights reserved.
doi:10.1016/j.mssp.2004.09.127

technological superiority into a sizeable market share. Aim of this paper is to present the current status of non-volatile memory technology, focusing on the perspective for Flash technology scaling and on potentialities and challenges of the emerging NVM concepts.

The current NVM mainstream is based on the Flash technology and it is expected that NOR and NAND Flash will be the high volume NVM production for the rest of this decade. To address the requirements that an NVM emerging concept must satisfy, NOR and NAND Flash performance, capabilities and limitations will be thus presented and discussed in Section 1. In Section 2, the attention will be focused on the innovative concepts for alternative NVM that have been proposed in the past and are under investigation today. Starting from already established concepts like ferroelectric (FeRAM) and magnetoesistive memories (MRAM), the fast growing phase-change memory (PCM) technology will be discussed. The ultimate memory concept provided by the cross-point zero-transistor approach, mainly based on polymeric materials, will be considered as they give an insight in next decade NVM concepts beyond the silicon-transistor era.

PCMs are one of the most promising candidates for next-generation NVM. Based on the reversible structural changes of chalcogenide materials, a fast write and read, good read signal window, very high endurance, and an intrinsic scalability are expected. The integration of a compact PCM cell structure and of the chalcogenide materials and the full compatibility with an advanced CMOS technology are key aspects to be demonstrated together with the optimization of the programming current [3,4]. In section 3 the current status of PCM technology will be reviewed, focusing on the recent developments employed to reduce the programming current. PCM technology scaling perspectives and rules will be finally addressed.

2. Flash technology

Recent technological developments have definitely confirmed the division of the NVM market into two mainstreams based on different applications of the Flash technology concept. By exploiting its multi-level capabilities (MLC) and the fast random access for execution in place applications, NOR memories are a cost effective solution to stack code and data storage into a single chip. With a cell size close to the minimum litho defined feature of $4F^2$ and MLC capabilities, NAND technology is expected to be the lowest cost NVM and the best solution for high-density data storage for the rest of this decade. The current visibility allows to predict a scenario where the floating-gate concept is a valuable solution till the 45 nm technology node (Fig. 1). From then on, the striking trade-off between oxide thickness and data

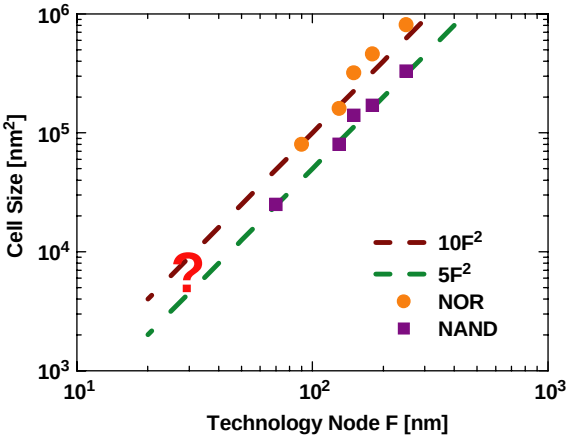

Fig. 1. Scaling trends for Flash technology.

retention should be faced with alternative solutions. Nano-floating gate (or nanocrystal) memories have been considered as the potential evolutive approach to overcome this limitation and further scale down the conventional Flash cell architecture [5,6]. The envisioned nanocrystal memory has a CMOS-compatible floating-gate-like cell structure. Instead of injecting charge in the floating gate, electrons are trapped into silicon nanocrystals that act as nano-floating gates (Fig. 2). By using electrically isolated charge-storage silicon dots, charge leakage through localized oxide defects is greatly reduced. This can be used to build floating gate structures with much thinner tunneling oxide layers. A major benefit of the nano-floating gate approach is the improved reliability. An endurance of 10^6 cycles has been demonstrated, slightly higher than the traditional floating gate approach. This approach is fully compatible with the standard CMOS technology, but some concerns still exist about the low threshold voltage shift (<3 V), data retention capabilities, and the intrinsic scalability of nanocrystals. Although the floating-gate concept has provided enthusiastic results, none of the existing and forecasted Flash memory architecture is able to optimize the entire performance set, which would bring it close to the universal memory specifications. Beyond nanocrystals, to enlarge application segments, offering better performance and scalability, new materials and alternative memory concepts are mandatory to boost the NVM industry.

3. Emerging non-volatile memories

To improve the performance and scalability of the floating gate devices, innovative concepts for alternative NVM have been proposed in the past and are under investigation today, dreaming to find the ideal memory that combines fast read, fast write, non-volatility,

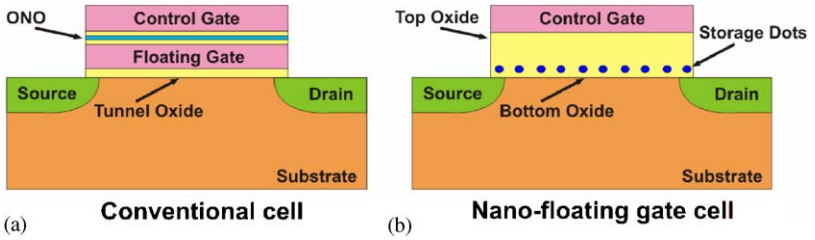

Fig. 2. Schematic structure of a conventional Flash cell and a nano-floating gate memory.

low-power, unlimited endurance and obviously at a cost comparable with Flash or DRAM. Two main categories of emerging NVM technologies have been so far investigated, one based on inorganic materials, like FeRAM, MRAM or PCM, and another based on organic materials, like ferroelectric or conductance switching polymers.

FeRAM is the only alternative NVM that has been commercialized so far, even if several challenging technology problems, mainly related to new materials and new manufacturing technologies, are still limiting their diffusion. Two classes of ferroelectric materials are currently used for FeRAM memories: perovskite structures and layered structures. Actually, the most used perovskite material for ferroelectric memories is $PbTi_xZr_{1-x}O_3$, also called PZT, while the layered ferroelectric choice for FeRAM memories is a Strontium–Bismuth–Tantalite alloy, which is referred to as SBT. Ferroelectric materials can be polarized spontaneously by an electric field. The polarization occurs as a lattice deformation of the cubic form below the Curie point, the temperature above which the material becomes paraelectric. For example, in PZT the Ti atom can be moved by an electric field into two stable positions, that are above and below the oxygen plane of the structure.

Endurance, also called electric fatigue, is the most important reliability characteristic and it is related to the ability decrease to switch the memory cell into the opposite state, after being kept programmed in one state for long periods of time. This effect is related to the polarization shift in the hysteresis loop and it is proportional with the increasing number of switching cycles. Nevertheless, the write cycle (but also the read cycle, since several proposed cell structures have a destructive reading) is expected to have an endurance level of about 10^{12}, which will suffice for a wide majority of high-demanding storage applications. Up to now several FeRAM cell structures have been proposed, with the ferroelectric material integrated either into a separate storage element, i.e. a ferroelectric capacitor (Fig. 3) [7], or into the selection element, i.e. a Ferroelectric FET transistor [8]. In the latter case, the storage and the selection elements are merged. The first cell type can be used in both the two-transistor/two-capacitor (2T2C) cell and the one-transistor/one-capacitor (1T1C) cell, while the later has been proposed with a one-transistor (1T) approach. All FeRAM architectures have high speed access cycles and provide genuine random access to all memory locations. Among the proposed architectures, the 1T1C FeRAM approach provides a cell size comparable to that of DRAMs with a simple on-chip peripheral circuits. The 1T FeRAM architecture is instead the most promising with a very small cell size and practically an infinite endurance level, but its processing has proven to be very complicated.

All current development efforts for MRAM technology are MTJ cell based (Fig. 4) [9] with an architecture composed by one transistor and one resistor (1T/1R). These technologies relies on the adoption of a tunnel junction coupled to magnetoresitive materials that exhibit changes in the electric resistance when a magnetic field is applied. The non-destructive read with a very fast access cycle is the premise for high performance, equal-long read and write cycles and for low power operation. Moreover, the structure is radiation hard with an unlimited read/write endurance, which makes MRAMs suitable for write intensive storage applications. The major MRAM disadvantage appears to be the high write current. While this technology has enough read current to guarantee a fast access time, it requires a very large write current (mA range), which increase power consumption. Despite FeRAM and MRAM NVM cells are very promising for high performance and low power systems, to date their development has faced many challenges that impact the manufacturability of high-density devices.

The PCM technology is one of the best candidate to cover different NVM application fields, matching both the high density as well as the high performance specifications. PCM exploits thermally reversible phase transitions of chalcogenide alloys (e.g., $Ge_2Sb_2Te_5$). The basic cell structure is composed by one transistor and one resistor (1T/1R) that can be programmed through the current induced Joule heating, and can be read by sensing the resistance change between the amorphous and the poly crystalline phase (Fig. 5). Despite this technology dates back to the '70 s, its application for

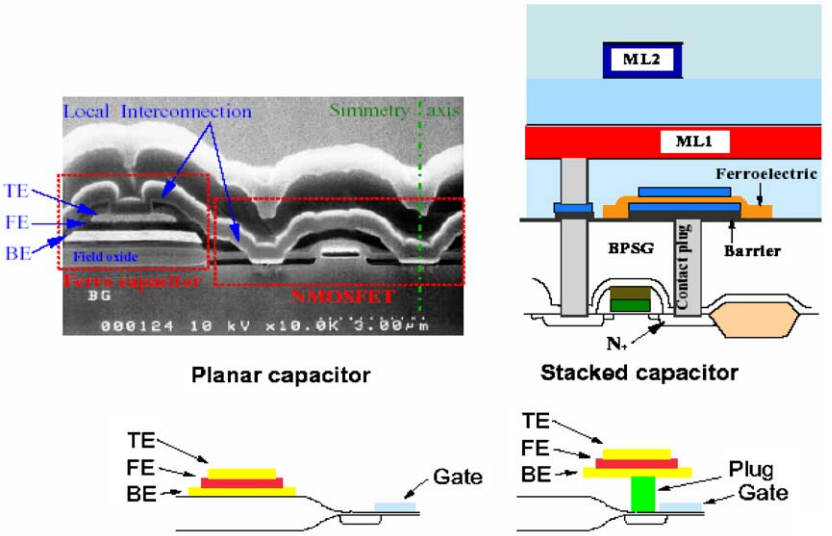

Fig. 3. Schematic structure of an FeRAM cell with the stacked capacitor approach.

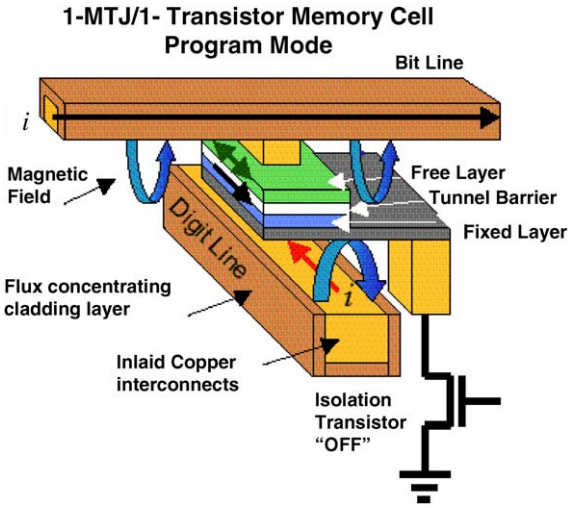

Fig. 4. MRAM cell in the 1MTJ/1 Transistor option [9], schematically showing the programming operation mode.

NVM has known renewed efforts in these years [10,11], showing the capabilities to reach the maturity for manufacturability [12,13].

Apart for the intrinsic scaling limits of each technology reported in Table 1, to date a common constrain to miniaturize an NVM cell beyond the 32 nm node is constituted by the selector transistor scaling. To address this issue, several cross point ($4F^2$) zero-transistor cells have been so far proposed, as the 0T/1R threshold switching PCM cell (TF-RRAM) [14], the polymeric ferroelectric RAM (PFRAM) [15], or the polymeric

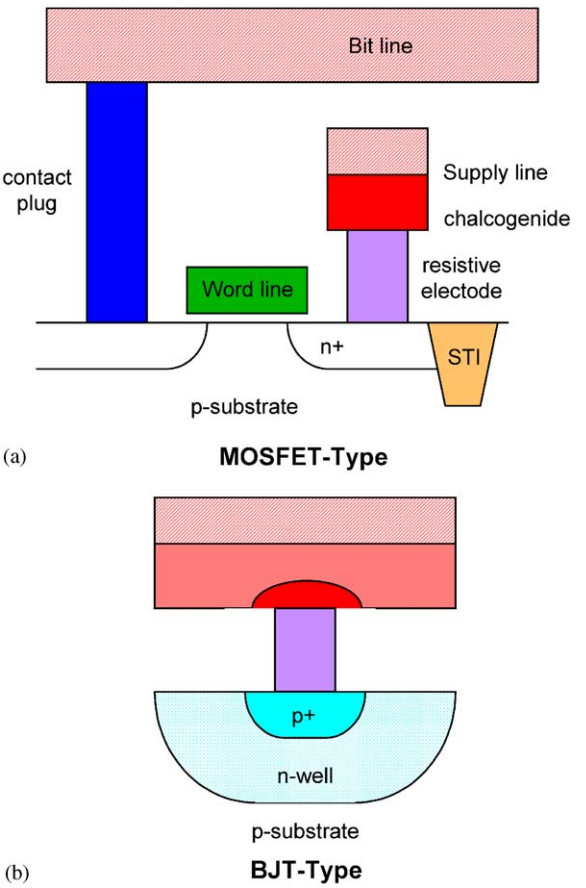

Fig. 5. PCM cell schematic structure with (a) MOSFET and (b) BJT selector.

Table 1
Comparison between Flash technology and the emerging NVM alternatives

	Flash		FeRAM	MRAM	PCM	Polymer	
Cell type	NOR 1T	NAND 1T	1T/1C	1T/1R	1T/1R	Ferro. 0T/1C	Res. 0T/1R
Cell size (F^2)	10	5	30–100	10–30	8–10	4	4
Endurance write/read	$10^6/\infty$		$10^{12}/10^{12}$	$>10^{14}/\infty$	$>10^{12}/\infty$	$10^9/10^9$	?
Read time (random)	60 ns	60 ns/serial	80 + 80 ns	30 ns	60 ns	10 μs	?
Write time (byte)	1 μs	200 μs/page	(read + write	30 ns	10 ns	(read + write	?
Erase time (byte)	1 s/sector	2 ms/block	destructive read)	30 ns	100 ns	destructive read)	?
Scalability	Fair		Poor	Poor	Good	Good	
Scalability limits	Tunnel oxide, HV		Capacitor	Current density	Litho	Litho	
Multi-bit capability	Yes		No	No	Yes	No	
3D Potential	No		No	No	Yes	Yes	
Relative Cost x bit	Medium	Low	High	High	Medium	Low	
Maturity	Very high		Low	Low	Low	Very low	

resistive memories [16]. Although several concerns still exist on the performance, feasibility, and manufacturability, the cross-point NVM based on a zero-transistor cell is considered as the ideal solution to continue to follow Moore's law, combining great interests for the research challenges and for the industrial potentialities.

4. Phase-change memories

Phase-change memories, also called Ovonic Unified Memory, are one of the most promising candidates for next-generation NVM [3,11,13,17], having the potentiality to improve the performance compared to Flash—random access time and read throughput (versus NOR), write throughput (versus NAND), direct write, bit granularity, endurance—as well as to be scalable beyond Flash technology. The PCM cell is essentially a resistor of a thin-film chalcogenide material ($Ge_2Sb_2Te_5$-GST) with a low-field resistance that changes by orders of magnitudes, depending on the phase state of the GST in the active region (i.e., crystalline or amorphous). In memory operation, cell read out is performed at low bias. Programming requires instead a relatively large current, in order to heat-up the GST and lead to a thermally induced local phase change. Phase-transitions can be thus easily achieved by applying voltage pulses with different amplitudes and with durations in the range of tenths of nanoseconds [10,11].

Although the PCM concept has been widely demonstrated, the integration of chalcogenide materials in a CMOS technology is still under investigation. Up to now the feasibility assessment of large PCM arrays has been focused on the integration with the CMOS technology [10,13], the reduction of the power consumption [4,11] and the scaling perspectives [3,17].

The programming current is typically larger than 1 mA based on 180 nm lithography [10]. An innovative idea to reduce the programming current is the edge contact approach reported by Ha et al. [11]. The contact area is determined by a thin film thickness in one dimension and programming current of hundreds of μA have been achieved. A novel μ-trench device structure with currents in the range of 600 μA has also been recently claimed [11,13]. This structure is characterized by a small contact area and a vertically integrated architecture, thus allowing a BJT-selected PCM cell of $12F^2$ in a 0.18 μm CMOS technology. Starting from research works on rewritable optical memory disks, another proposal to reduce the programming current is to dope the chalcogenide material with nitrogen [18]. Higher resistance material is achieved by nitrogen doping and this leads to reduced programming currents.

To exploit the PCM technology for high-performance applications, fast write and read time are mandatory, still preserving good data retention capabilities. Fig. 6 shows the programming curves of a PCM cell. For pulses as low as 80 (SET) and 20 ns (RESET) a resistance change of two order of magnitude between the programmed states can be easily achieved. One of the main concern to speed up the writing operation is the trade-off condition between fast crystallization and good non-volatility properties. Retention capabilities have been thus assessed and an activation energy of 2.6 eV has been measured. Fig. 7 shows that the RESET state can be retained at 110 °C 10 years, while data retention of more than 300 years can be estimated at 85 °C. An intrinsic endurance longer than 10^{11} program-

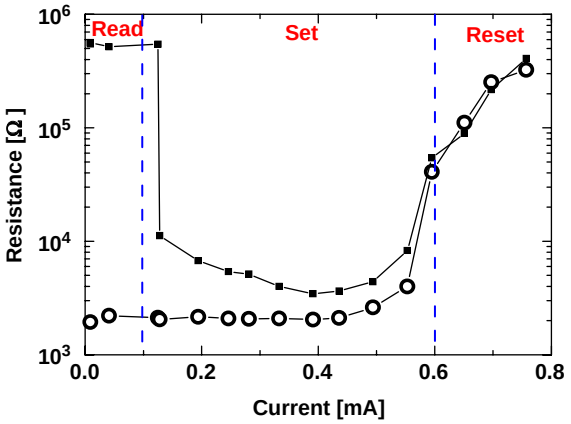

Fig. 6. Programming curves of a PCM cell.

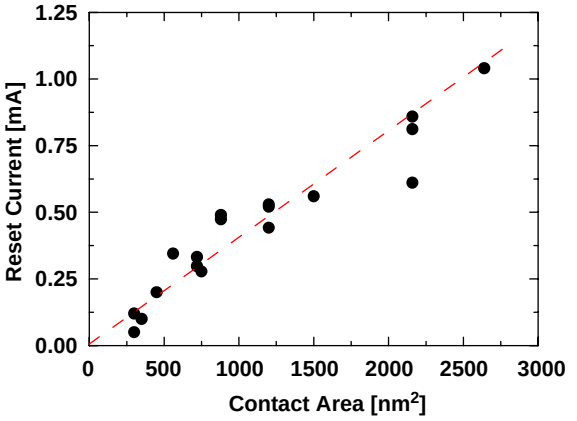

Fig. 8. Experimental reset current as a function of the contact area. RESET currents as low as 50 μA can be obtained.

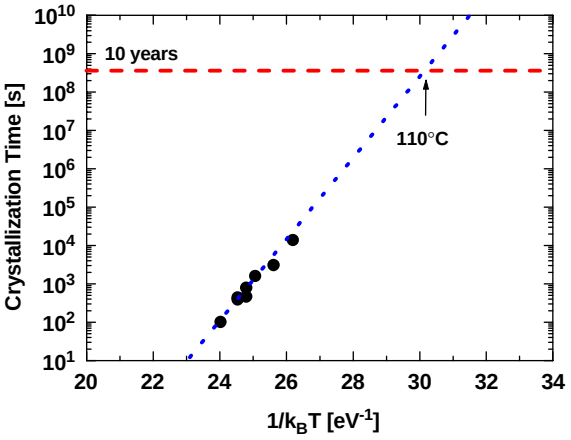

Fig. 7. Data retention measured on a PCM cell.

ming cycles has also been claimed [10,11], confirming the high potentialities of PCM technology.

One of the main concern related to the PCM technology relies on the impact of the thermal cross-talk between adjacent bits. With the perspective to scale down the device dimension, the distance between cells in a high-density memory device is of the order of the minimum lithographic pitch. If a high current pulse is applied, the neighbor bits can reach the crystallization temperature, thus losing the stored information. A numerical model has been used to investigate the scaling potential of PCM technology [17]. Even in worst case conditions, numerical results suggest a negligible thermal disturb between adjacent bits in both transient and steady-state condition, confirming the high scaling capabilities of the PCM technology beyond the 45 nm technology node.

The scalability of the reset current have been experimentally addressed measuring several test devices with different contact area. The resulting values are reported in Fig. 8. From this figure it is clear that the RESET current follows the reduction of the contact area and values as low as 50 μA have been achieved with a complete device functionality. It has also been shown [17] that the RESET current strongly depends on the device geometry and materials, thus suggesting that a further power reduction may be achieved by a judicious device engineering.

Table 2 summarizes instead the isotropic scaling rules expected if each geometric dimension is scaled by a factor $k>1$. This results into a contact area scaling as $1/k^2$ and a heater/GST layer thickness scaling as $1/k$. Since both the electrical and thermal resistances increase linearly with k, the reset current is expected to scale as $1/k$. Analogously, the current capability of the device used as cell selector (BJT or MOSFET) is expected to scale linearly [1], thus matching the requirements for PCM current scaling. However, the constant voltage drop on the PCM device reported in Table 2 could be a concern to scale the PCM technology. Considering an array structure with MOSFET or BJT selectors, such voltage could become an issue for the gate oxide reliability of unselected scaled MOSFET devices, while the breakdown voltages of unselected BJT should be enough to sustain the applied voltage. An accurate design of the PCM selecting device will thus become mandatory to scale down this technology beyond the 45 nm technology node.

5. Conclusions

The current NVM mainstream is based on the Flash technology and it is expected that NOR and NAND Flash will be the high volume NVM production for the rest of this decade. Although the floating-gate concept

Table 2
PCM scaling rules

	Parameters	Factor ($k>1$)
PCM cell	GST/heater contact area	$1/k^2$
	GST thickness	$1/k$
	Heater height/thickness	$1/k$
	SET resistance	k
	RESET resistance	k
	Thermal resistance	k
	ON-state resistance	k
	Threshold voltage	$\log(k)$
	Programming voltage	1
	Programming current	$1/k$
BJT [1]	Emitter area	$1/k^2$
	Base thickness	$1/k$
	Emitter current	$1/k$
MOS [1]	Gate length	$1/k$
	Width	$1/k$
	Saturation current	$1/k$

has provided enthusiastic results, none of the existing and forecasted Flash memory architecture is able to optimize the entire performance set, which would bring it close to the universal memory specifications. To enlarge application segments, offering better performance and scalability, new materials and alternative memory concepts are mandatory to boost the NVM industry. Among the innovative concepts for alternative NVM that have been proposed in the past and are under investigation today, phase-change memory appears as one of the best candidate to cover different NVM application fields, matching both the high density as well as the high performance specifications.

References

[1] ITRS. Updated Edition, 2003.
[2] Bez R, Camerlenghi E, Modelli A, Visconti A. Proc IEEE 2003;91(4):489.
[3] Lai S. IEDM Tech. Dig. 2003;255.
[4] Hwang YN, Lee SH, Ahn SJ, Lee SY, Ryoo KC, Hong HS, Koo HC, Yeung F, Oh JH, Kim HJ, Jeong WC, Park JH, Horii H, Ha YH, Yi JH, Koh GH, Jeong GT, Jeong HS, Kinam Kim. IEDM Tech. Dig. 2003;893.
[5] Muralidhar R, Steimle RF, Sadd M, Rao R, Swift CT, Prinz EJ, Yater J, Grieve L, Harber K, Hradsky B, Straub S, Acred B, Paulson W, Chen W, Parker L, Anderson SGH, Rossow M, Merchant T, Paransky M, Huynh T, Hadad D, Ko-Min Chang, White BE Jr. IEDM Tech. Dig. 2003;601.
[6] De Salvo B, Gerardi C, Lombardo S, Baron T, Perniola L, Mariolle D, Mur P, Toffoli A, Gely M, Semeria MN, Deleonibus S, Ammendola G, Ancarani V, Melanotte M, Bez R, Baldi L, Corso D, Crupi I, Puglisi RA, Nicotra G, Rimini E, Mazen F, Ghibaudo G, Pananakakis G, Compagnoni CM, Ielmini D, Lacaita A, Spinelli A, Wan, YM, van der Jeugd K. IEDM Tech. Dig. 2003;597.
[7] Oh S-H, Hong, S-K, Kweon S-Y, Kim N-K, Yang Y-H, Kim J-G, Seong J-Y, Jang I-W, Park S-H, Bang K-H, Lee K-N, Jeong H-J, Son J-H, Lee S-S, Choi E-S, Sun H-J, Yeom S-J, Ban K-D, Park J-W, Park G-D, Song S-Y, Shin J-H, Lee S-I, Park Y-J. IEDM Tech. Dig. 2003;835.
[8] Ishiwara H. IEDM Tech. Dig. 2003;263.
[9] Durlam M, Addie D, Akerman J, Butcher B, Brown P, Chan J, DeHerrera M, Engel BN, Feil B, Grynkewich G, Janesky J, Johnson M, Kyler K, Molla J, Martin J, Nagel K, Ren J, Rizzo ND, Rodriguez, T, Savtchenko L, Salter J, Slaugher JM, Smith K, Sun JJ, Lien M, Papworth K, Shah P, Qin W, Williams R, Wise L, Tehrani S. IEDM Tech. Dig. 2003;995.
[10] Lai S, Lowrey T, IEDM Tech. Dig. 2001;803.
[11] Pellizzer F, Pirovano A, Ottogalli F, Magistretti M, Scaravaggi M, Zuliani P, Tosi M, Benvenuti A, Besana P, Cadeo S, Marangon T, Morandi R, Piva R, Spandre A, Zonca R, Modelli A, Varesi E, Lowrey T, Lacaita A, Casagrande G, Cappelletti P, Bez R. 2004 VLSI Symp. on Tech. 2004.
[12] Woo Yeong Cho, Cho B-H, Choi B-G, Oh H-R, Kang S-B, Kim K-S, Kim K-H, Kim D-E, Kwak C-K, Byun H-G, Hwang Y-N, Ahn S-J, Jung G-T, Jung H-S, Kinam Kim. ISSCC 2004, 2004;40.
[13] Bedeschi F, Resta C, Khouri O, Buda E, Costa L, Ferraro M, Pellizzer F, Ottogalli F, Pirovano A, Tosi M, Bez R, Gastaldi R, Casagrande G. 2004 VLSI Symp. on Circ. 2004.
[14] Yi-Chou Chen, Chen CF, Chen CT, Yu JY, Wu S, Lung SL, Rich Liu, Chih-Yuan Lu. 2003 IEDM Tech. Dig. 2003;905.
[15] Web site: http://www.thinfilm.se.
[16] Sven Möller, Craig Perlov, Warren Jackson, Carl Taussig and Stephen R. Forrest. Nature, 2003;426:166.
[17] Pirovano A, Lacaita AL, Benvenuti A, Pellizzer F, Hudgens S, Bez R. IEDM Tech. Dig. 2003;699.
[18] Ha YH, Yi JH, Horii H, Park JH, Joo SH, Park SO, U-In Chung, Moon JT. Symp. VLSI Tech. 2003.