Overview of candidate device technologies for storage-class memory

G. W. Burr
B. N. Kurdi
J. C. Scott
C. H. Lam
K. Gopalakrishnan
R. S. Shenoy

Storage-class memory (SCM) combines the benefits of a solid-state memory, such as high performance and robustness, with the archival capabilities and low cost of conventional hard-disk magnetic storage. Such a device would require a solid-state nonvolatile memory technology that could be manufactured at an extremely high effective areal density using some combination of sublithographic patterning techniques, multiple bits per cell, and multiple layers of devices. We review the candidate solid-state nonvolatile memory technologies that potentially could be used to construct such an SCM. We discuss evolutionary extensions of conventional flash memory, such as SONOS (silicon-oxide-nitride-oxide-silicon) and nanotraps, as well as a number of revolutionary new memory technologies. We review the capabilities of ferroelectric, magnetic, phase-change, and resistive random-access memories, including perovskites and solid electrolytes, and finally organic and polymeric memory. The potential for practical scaling to ultrahigh effective areal density for each of these candidate technologies is then compared.

Introduction

As originally formulated by Gordon Moore in 1965 [1], Moore's Law is fairly simple: a prediction that the number of devices that can be integrated on a chip of fixed area would double every 12 months. This simple prediction (later amended to doubling every 18–24 months) unleashed a powerful economic cycle of investment followed by enhanced products followed by new and varied applications motivating yet more investment. Thus, Moore's Law has become the driving force behind dramatic reductions in unit cost over the past few decades for memory, enabling products of ever higher density and ultimately putting enormous amounts of memory in the hands of the consumer at much reduced cost. For example, the cost of flash memory has fallen from $600 per megabyte in 1987 to $0.01 per megabyte in 2007 (a factor of 60,000 in 20 years, corresponding to halving the cost by doubling the density every 15 months) (**Figure 1**).

This powerful economic cycle has made the prediction of near-future product developments amazingly reliable because the underlying device physics, materials, and fabrication processes have all been scalable, at least until now [2]. However, beyond the end of this decade it will be hard to continue to shrink the ubiquitous nonvolatile memory (NVM) known as *flash memory* [3, 4].

Such a breakpoint presents a great opportunity for alternative technologies. However, in order to replace flash, an alternative technology will have to be superior to it in some combination of such factors as further scalability, cost per bit, and performance (e.g., memory speed). However, the size of the opportunity goes well beyond simply providing a potential successor for flash technology despite the many applications that it currently addresses. In fact, the emergence of an NVM solid-state memory technology that combines high performance, high density, and low cost could usher in seminal changes in the memory and storage hierarchy for all computing platforms ranging up to high-performance computing. If the cost per bit could be driven low enough through ultrahigh memory density, ultimately such a storage-class

©Copyright 2008 by International Business Machines Corporation. Copying in printed form for private use is permitted without payment of royalty provided that (1) each reproduction is done without alteration and (2) the *Journal* reference and IBM copyright notice are included on the first page. The title and abstract, but no other portions, of this paper may be copied by any means or distributed royalty free without further permission by computer-based and other information-service systems. Permission to *republish* any other portion of this paper must be obtained from the Editor.

0018-8646/08/$5.00 © 2008 IBM

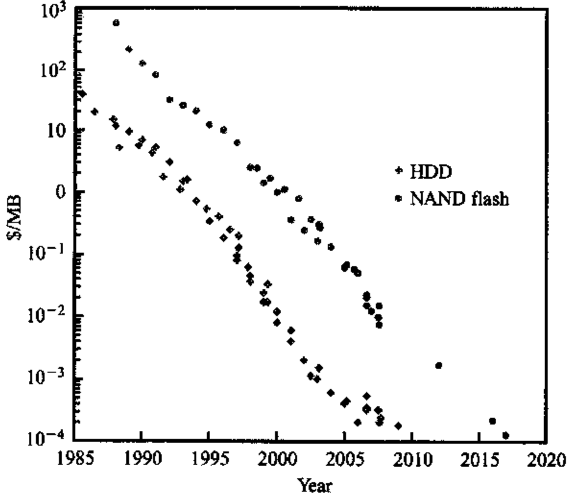

Figure 1

Cost per megabyte of desktop magnetic hard disk drives (HDDs) and NAND flash, compiled from various sources, including Gartner Dataquest and periodic web searches for consumer prices.

memory (SCM) device could potentially replace magnetic hard-disk drives (HDDs) in enterprise storage server systems.

In this paper, we briefly discuss possible low-cost integration approaches and then review the candidate NVM solid-state memory technologies that potentially could be used to construct such an SCM.

SCM

It should be noted that Moore's Law addresses only the density of an integrated circuit, predicting the fabrication of semiconductors with smaller features, which in turn drives the economics. The classical scaling of devices is more formally addressed by Dennard's Law, which calls for the coordinated miniaturization of a small set of device parameters, which together dictate overall performance [5]. Both laws need to be considered because when classic scaling no longer produces the hoped-for performance improvements, the economic benefits typically associated with Moore's Law may no longer follow.

Historically, when scaling both drove cost down and increased performance (in other words, Moore's Law and Dennard's Law were synchronous), an unrelenting focus on processor performance and the scaling of logic devices was justified. However, as the processor has become power constrained and the scaling of logic devices no longer results in direct performance improvements, it is natural to consider developing additional avenues for improvement. For instance, one might consider how to get more overall system performance from changes in the memory-storage hierarchy. The best-case access time of a magnetic HDD has remained fairly constant over the past decade at approximately 3–5 milliseconds [6]. A 3-GHz microprocessor will execute nine million instructions while waiting for this data. In fact, an enormous part of a modern computing system is designed expressly to hide or finesse this unpleasant discrepancy in performance [7].

Because critical computing applications are becoming more data-centric (as in *mine this database*) than compute-centric (as in *solve this differential equation*) [8], a high-performance, high-density, and low-cost NVM technology whose access time falls between that of an HDD and the dynamic memory located near the processor would significantly improve overall system performance. Such a memory technology would be a welcome near-future development.

There has also been a somewhat similar cycle of miniaturization at work in the magnetic HDD industry. An enormous growth in areal density (a 35-million-fold increase from 1957 to 2007 driven by a 100% annual compound growth rate) has produced dramatic reductions in the cost per bit of magnetic storage [6]. This has made it attractive to solve storage system performance issues by simply adding disk drives, as shown by the use of redundant disk arrays to compensate for bandwidth and latency limitations in enterprise storage servers. However, in the year 2020 this trend will call for millions of HDDs in large server installations [8]. In such a situation, these disks not only consume most of the overall space and power budget [8], but the logistics of dealing with failures during recovery from a drive failure become extremely difficult. These issues can no longer be managed by simply adding more drives, no matter how low their cost.

In summary, on one side there is the goal to develop a nonvolatile, low-cost, high-performance, solid-state memory that could extend beyond flash memory; on the other side is the need for a solid-state memory technology to meet the demands of future storage server systems. Bridging the two is the potential for significant system performance improvements in all types of computing systems with the insertion of such an NVM technology within the storage-memory hierarchy. Given the powerful forces in search of the target performance specifications given in **Table 1**, we believe that some type of new SCM technology is quite likely to emerge.

Because the expected ultimate performance of currently available flash technology will likely fall short of these requirements, here we examine the potential of other emerging NVM technologies. However, we emphasize

that the density and cost requirements of SCM transcend the straightforward application of Moore's Law. For instance, currently the most inexpensive semiconductor memory technology is multilevel cell (MLC) NAND flash because it offers the highest effective areal density of 1 stored bit per 2 F^2 of silicon substrate area, where F is defined as the smallest lithographic dimension in a given technology node. This effective density combines the benefits of following Moore's Law (the smallest lithographically defined memory cell of 4 F^2) together with MLC technology providing 2 bits per memory cell. Some of the insufficient performance of flash in terms of access time and endurance could potentially be finessed by a hybrid system approach, such as a dynamic RAM (DRAM) cache. However, even if we could ignore the costs incurred by such an approach, the MLC NAND flash alone would still fail to meet the future SCM density and cost requirements shown in Table 1.

Thus, other techniques will need to be invoked in order to achieve the ultrahigh memory densities demanded by SCM. These can be grouped into three possibilities:

1. Three-dimensional (3D) integration of multiple layers of memory, currently implemented commercially for write-once solid-state memory [9].
2. Multiple bits per cell using MLC techniques [10].
3. Sublithographic crossbar memory to go beyond the lithographic dimension, F [11].

It is beyond the scope of this paper to address these approaches, not only because they are well covered in the mentioned references, but because the implementation of techniques such as these that go beyond 4 F^2 will be critically dependent on the choice of memory device. However, as we focus our attention on emerging NVM device technologies, it is important to keep in mind that achieving low cost through ultrahigh density while maintaining memory performance (nonvolatile retention, high endurance, high read and write bandwidths, and fast access time) will be absolutely crucial to the success of SCM.

We begin with the established technology of flash memory and move to newer charge-based variants such as SONOS (silicon-oxide-nitride-oxide-silicon) memory and ferroelectric RAM. We then briefly review magnetic RAM, and then turn to phase-change memory, resistive RAM, solid-electrolyte, and organic memory. These technologies have received strong industry interest, as indicated in **Figure 2**, which shows papers presented in these areas over the past 7 years at two major conferences: the Symposium on VLSI Technology and the International Electron Devices Meeting (IEDM).

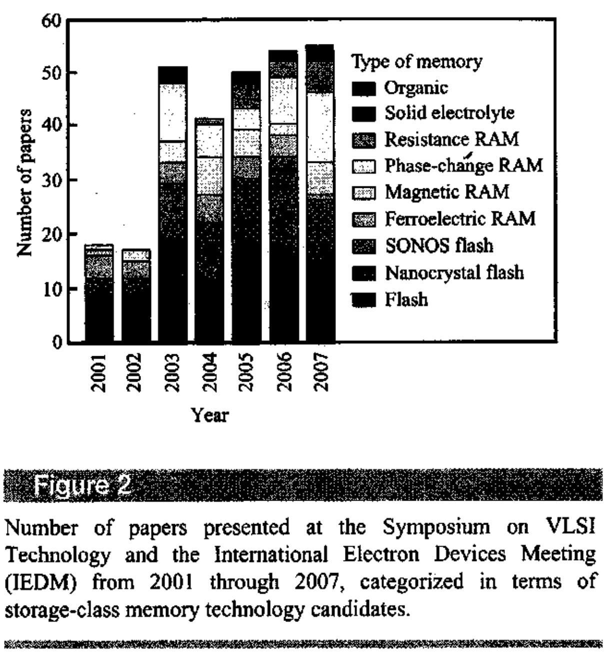

Figure 2

Number of papers presented at the Symposium on VLSI Technology and the International Electron Devices Meeting (IEDM) from 2001 through 2007, categorized in terms of storage-class memory technology candidates.

Table 1 Target specification for SCM: a combination of the reliability, fast access, and endurance of a solid-state memory, together with the low-cost archival capabilities and vast capacity of a magnetic hard disk drive.

Access time	50–1,000 ns
Data rate	100 MB/s
Endurance (cycles)	10^9–10^{12}
Hard error rate (bits/terabyte)	10^{-4}
Mean time between failures	2 million hours
Data retention	10 years
On power	100 mW
Standby power	1 mW
Cost	≪$5/GB
Annual compound growth rate	35%

Flash and other charge-trapping memory

In a conventional metal-oxide semiconductor transistor, a voltage applied to the gate allows current to flow from the source to the drain. To implement either floating-gate or charge-trapping one-transistor memory, the gate has been redesigned to allow electrons to be placed (or removed) near the gate during a writing step. The presence (absence) of this charge shifts (restores) the threshold voltage of the transistor, allowing detection of the binary state of the memory cell. The floating-gate memory became the preferred device because the ease of erasing stored charge enabled a memory device that was both

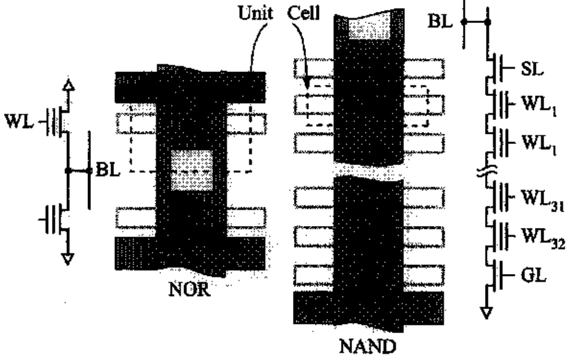

Figure 3

Circuit schematics and top-down memory cell layouts for conventional NOR and NAND flash memory. (BL: bitline; GL: ground select line; SL: select line; WL: wordline.)

nonvolatile and reprogrammable. Early floating-gate devices were erased with a few minutes of ultraviolet radiation, which imparted enough energy for stored electrons to surmount the insulating barrier. In a modern floating-gate memory, whole blocks of devices are erased electrically in less than a second, giving rise to the term *flash memory* [12]. This change greatly simplified memory packaging and simultaneously opened up a vast new set of applications [4, 12].

Today, there are two major kinds of flash memories, NOR and NAND, as depicted in **Figure 3** [4, 12]. In NOR memory, each cell in a memory array is directly connected to the wordlines and bitlines of the memory array, while NAND memory devices are arranged in series within small blocks. Thus, while NAND flash can inherently be packed more densely than NOR flash, NOR flash offers significantly faster random access. For example, the read bandwidth for NOR can exceed 100 MB/s, while NAND ranges from 18 to 25 MB/s [13]. Because NOR flash memory is programmed using channel hot-electron injection (CHEI) and erased using Fowler–Nordheim tunneling [12], both writing and erasing are extremely slow (0.18–0.47 MB/s and 750–900 ms, respectively [13, 14]). With the combination of fast readout and slow programming, NOR is being used in applications in which it can serve as a programmable read-only memory that offers fast access to data that is modified only occasionally. For instance, program code in many cell phones, computers, and other devices is now loaded directly from NOR flash, representing an estimated $8 billion annual market.

In contrast to NOR cells programmed with CHEI, Fowler–Nordheim tunneling is used to program NAND flash memory through the channel area. While this is inherently quite slow per bit, the very low current required for this process allows many bits to be written in parallel, resulting in a reasonable write bandwidth (8 MB/s) [13]. The introduction of NAND flash allowed a reduction in cell size from the 9–11 F^2 still used for NOR cells down to roughly 4 F^2 [2]. Another enabling technology was the MLC concept [10] in which each memory cell stores 2 bits of information (making the effective cell size per bit 2 F^2). Advanced MLC with up to 4 bits per cell is being developed but as of this writing is not yet in production [2].

Because of the block-based architecture, the random access time to any given bit tends to be slow (25 μs), with a significantly faster readout of data blocks (23–37 MB/s) after this initial access delay [13, 14]. Erasing a block tends to be extremely slow (2 ms) [13, 14]. This slow random-access performance means that NAND flash is best suited for applications that primarily require block-based access, such as the storage of digital music, photos, and video. Despite these limitations, the popularity of these applications has resulted in the growth of the NAND flash market to $14.2 billion for 2007 [15]. In 2007, it was estimated that about 1.6×10^{18} bytes of NAND flash will be ship—more capacity in 1 year than DRAM chips have shipped since the first commercial one was sold in 1972 [16].

So far, the basic floating-gate device has served the flash industry well. In addition to known challenges common throughout the semiconductor industry related to shrinking the lithography pitch and the increasing importance of device-to-device variations in upcoming technology generations, there are a few unique technical challenges in scaling the floating-gate NAND flash memory beyond 40 nm [2–4, 17]. Stringent data-retention requirements, particularly for MLC, set a practical limit for the thickness of the tunnel oxide of roughly 7 nm [2].

Another challenge in the scaling of floating-gate NAND is floating-gate interference [18]. For effective programming using tunneling, the gate aspect ratio must stay relatively constant in order to maintain the coupling between the control gate and the floating gate [12]. However, this creates unacceptable interference between adjacent memory devices when the spacing between wordlines shrinks to 40 nm or less [3, 4, 17]. One obvious alternative is to replace the floating gate with a charge-trapping layer, such as the silicon nitride in the SONOS cell structure [19]. Early SONOS memory devices used extremely thin tunnel and blocking oxides for acceptable write and erase performance, and thus they suffered from issues with data retention [20]. Recent advances in metal gates and high-k dielectric materials research have provided improvements in erase and retention characteristics [21]. **Figure 4** depicts the structural differences between the floating gate, SONOS, and the

$TaN-Al_2O_3$-nitride-oxide-silicon (TANOS) [21–23] memory cell structures. Initial TANOS NAND memory cells demonstrated acceptable memory cell characteristics for binary operations [22], and recent developments with a bandgap-engineered charge-trapping layer in which an ultrathin sandwich of different materials is introduced to obtain low leakage yet low programming voltages have demonstrated reliable MLC operations with 2 bits per cell [23]. Another avenue is to intentionally introduce traps in the form of nanocrystals [24].

Thus, it is now generally believed that by moving to charge-trapping storage, NAND flash will be able to scale to at least the 22-nm technology generation [3, 17]. The current aggressive geometry migration calling for NAND flash to scale faster than the capabilities of photolithography can also be maintained by using self-aligned double patterning technology [25] to define memory cell arrays at ultrasmall pitch.

The challenge not only for NAND flash but for all charge-based memories at these dimensions is the limited number of stored electrons available for MLC operations. It has been shown that losing even 100 electrons could result in a retention failure for 2-bit MLC operation at the 40-nm generation in a 3D channel memory cell [3]. This charge-loss issue gets worse with continued scaling of the cell volume. Thus, to continue to shrink the cell footprint and decrease the cost while still maintaining the amount of charge being stored will require some type of extensive 3D integration scheme. A stacked surrounding-gate transistor NAND flash memory cell structure has been proposed [26] and successfully demonstrated using a 4-bit NAND SONOS memory cell structure [27]. Other 3D stacking integration schemes using polysilicon or thin-film transistors to construct the NAND flash have also been proposed [28].

In order to satisfy all of the specifications in Table 1 and thus take advantage of all the applications waiting for a true SCM, flash memory must improve in both endurance and performance while aggressively moving to ultrasmall device dimensions in spite of the known scaling issues.

Ferroelectric RAM

A ferroelectric capacitor is formed by sandwiching a ferroelectric material such as $Pb(Zr_xTi_{1-x})O$, lead zirconate titanate, also known as *PZT*, between two metallic electrodes. In the resulting hysteresis loop of charge as a function of voltage, the two stable states at zero applied voltage represent a remanent polarization produced by switching the spontaneous polarization of the material [29]. However, since ferroelectric materials tend to have soft, nonsquare hysteresis loops, the half-select operation in a crossbar memory array would perturb stored data on the nearby cells subjected to half

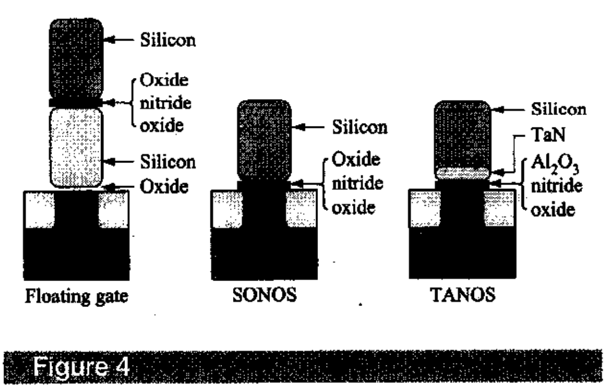

Figure 4

Cross-sectional view showing the thin-film layers composing the gate in conventional floating-gate, SONOS, and TANOS memory cells.

the programming voltages. This necessitates a selection device such as a transistor [29].

The most straightforward way to detect the state of such a ferroelectric capacitor is to apply a voltage pulse to take the device to one extreme of its hysteresis loop, producing a current spike whose magnitude depends on the initial state. The readout voltage produced by the charging of the bitline capacitance by this current can be compared with a reference voltage [29]. However, this readout technique is destructive because the device ends up in the same final state independent of the original data. Thus, each read access must be accompanied by a subsequent rewrite operation, which immediately pushes the required switching-cycle endurance to a very large number. For instance, even at a clock speed of 100 MHz, the worst-case scenario over a 10-year lifetime represents more than 10^{16} read cycles [30].

Despite this issue, ferroelectric RAM (FeRAM) was one of the strongest early candidates to be the next nonvolatile RAM because of its inherent speed (as low as 20 ns [2]), its low-power, low-voltage operation, and the possibility of straightforward CMOS (complementary metal-oxide semiconductor) integration [30]. Initially, FeRAM was seen as a way to simply have all the advantages of DRAM together with long-term nonvolatile storage [31]. However, despite years of concentrated integration efforts, FeRAM cells remain significantly larger than the DRAM cell size of 6–8 F^2 [31].

As with any well-explored technology, the problems are now well known. Because the output signal depends on transferring a charge 2 P_rA onto the bitline capacitance C_b to obtain a detectable voltage difference ΔV, the scaling of FeRAM to smaller device areas with smaller capacitor area A inherently leads to smaller signal levels [31]. Thus, the fabrication of FeRAM devices has moved from *strapped devices*, where the capacitor sits next to the

wordline, to *stacked devices*, where the capacitor sits directly above the wordline, to *3D devices* where the capacitor is conformally deposited in a smooth continuous layer either within a trench or over a ridge, augmenting the effective capacitor area without increasing the device footprint [31].

The ferroelectric capacitors in FeRAM devices also tend to show significant problems that include fatigue (the remanent polarization decreases with cycling), imprint (a device left in one state tends to favor this polarization, causing the hysteresis loop to shift), and retention (loss of stored polarization over time) [31–33]. Fatigue issues have been greatly reduced either by changing electrode materials away from metals such as Pt and Ir toward metal oxides such as RuO_2, IrO_2, and $SrRuO_3$ or by switching materials from PZT to either $SrBi_2Ta_2O_9$ (strontium bismuth tantalite, or *SBT*) or lanthanum-doped $Bi_{4-x}La_xTi_3O_{12}$ (bismuth titanate, or *BLT*) [31–33]. Unfortunately, changing ferroelectric materials tends to simply swap which aspects are strengths and weaknesses. PZT is known for high P_r and well-defined crystallization at moderate temperatures (~600°C) but can have significant imprint and retention issues; on the other hand, SBT offers improved reliability and lower coercive field but exhibits decreased P_r and poor control over crystallization orientation. SBT also tends to require higher processing temperatures (700–800°C) [31].

These high temperatures are particularly problematic given that during the processing, the metallic wordline just under the capacitor (as required for high-density memory) must be protected from being oxidized while the ferroelectric is deposited with a metal-organic chemical vapor deposition process [31]. Worse yet, the ferroelectric material must be protected at all times from hydrogen, either introduced by diffusion from the underlying wordline [34] or while forming the top electrode contact [35]. Defects associated with hydrogen diffusing into the ferroelectric capacitor have been shown to greatly accelerate imprint and fatigue problems [31, 34, 35]. Since the effective device P_r drops abruptly when the ferroelectric layer becomes thin (<100 nm) [31], presumably due to interface effects at the electrodes [33], even scaled-down devices are likely to require thick capacitors [31]. However, because it is very difficult to etch a metal-ferroelectric-metal stack without introducing a slope, the wide guard band needed to separate thick capacitors leaves at best only 50–70% of the effective device area for the capacitor itself [31].

A number of structures offering avenues for improvement have been proposed and developed. The 2-transistor–2-capacitor (2T–2C) concept offers twice as much signal and more reliable voltage referencing but at the significant cost of twice the device area [29]. The chain-FeRAM concept [36] assembles devices in series, like NAND flash, in order to decrease the bitline capacitance and thus increase the detectable signal, but this approach sacrifices access speed and offers only modest improvements in the effective areal density [29]. A very old concept [37] involves building a FeFET—a field-effect transistor with a ferroelectric capacitor—as the gate electrode in order to create an NVM element [32, 38]. Unfortunately, there are a number of problems with this including the necessity of integrating ferroelectric materials directly on silicon [38]. Also, if a dielectric capacitor is used to separate the ferroelectric from the silicon to ease integration, then the nonvolatile data lifetime decreases to mere weeks because of the inherent depolarizing field present in such a structure [32, 38]. A variant introduced to restore the nonvolatility is the 1T–2C concept, which avoids the depolarizing field by using two opposite-poled capacitors, but again at the cost of increased cell size [39].

In summary, most of the recent work on FeRAM seems to address embedded memory applications [40, 41] building on the strengths of FeRAM (CMOS compatibility, low-power, and low-voltage operation) and avoiding its weaknesses (difficulty in scaling to ultrasmall cell size). On the other hand, FeRAM is one of the most commercially successful new NVM alternatives, having been used in the Sony PlayStation** 2 system [32]. However, for FeRAM to be a viable alternative for SCM, there needs to be a significant breakthrough in the integration of ultrasmall cells using 3D ferroelectric capacitors without sacrificing reliability or memory performance.

Magnetic RAM

One of the key breakthroughs that drove the rapid improvement in the areal density of HDDs over the past 20 years [6] was the development of incredibly sensitive sensors for the detection of the weak magnetic fields associated with the data-bearing magnetic transitions on the disk [42]. Work on these sensors carried over to a closely related device, the magnetic tunnel junction (MTJ) [42]. The amount of spin-polarized tunneling current passing through the dielectric separating the two magnetic layers of an MTJ depends on the relative magnetizations of the two layers, an effect known as *tunneling magnetoresistance* (TMR) [42, 43]. One of the layers is designed to have its magnetization pinned, while the other is free to have its magnetization flipped by an external writing event. The pinned layer can be made more stable by using a pair of coupled ferromagnetic layers instead [42].

The advantages of such a magnetic RAM (MRAM) cell are an inherently fast write speed, straightforward placement above the silicon using the CMOS back end of

the line, and the prospect of very high endurance, as there is no known wear-out mechanism for magnetic switching [44]. In addition, the ability to write a cell by simply passing current through two nearby wires (causing the superimposed magnetic field to exceed the write threshold) would seem to enable a true cross-point memory composed of wires and MTJ devices.

However, small currents leaking through unselected devices in the array make it necessary to use a selection device such as a transistor [45]. In addition, although the absolute resistance of the cell could easily be tuned as desired by varying the tunnel oxide thickness [45], the change in resistance between the two states is fairly small compared to other NVM technologies.

More critically, the energy barrier that should prevent half-selected cells from switching tended to be particularly low [45] (*half-selected cells* are those that are influenced by the magnetic field from just one of the wires that leads to the cell being written). This problem would generate occasional bit errors in neighboring cells during the write process [45] and was solved by the toggle variant of MRAM [45], in which the free layer was also replaced by a coupled layer pair acting as a synthetic antiferromagnet [44]. This change made the device extremely insensitive to half-select perturbation but at the cost of requiring a read-before-write [45] because the write operation literally toggles the state of the bit.

However, the most serious problem for the use of MRAM, and one that remains a problem even with toggle MRAM, is that the write currents remain very high (>1 mA) and show no sign of decreasing as devices are scaled smaller [44]. These currents are sufficiently large—even in 180-nm technology—that electromigration-induced damage of the wires themselves has already become the major failure mechanism [44]. Thus, as with FeRAM, some fraction of the integrated device efforts has been refocused on the embedded or system-on-a-chip application [44].

A number of prospective MRAM variants that could help scale the technology down to ultrasmall dimensions have been proposed. Several proposals involve heating an MTJ device by passing a small current through the cell in order to reduce the threshold switching field and thus the amount of current on the nearby line [46]. Closely related are schemes in which the cell is directly switched by a current passing through the tunnel junction using the spin-torque effect [45, 47, 48]. Because this effect appears to depend on current density rather than current [48], the absolute current values should decrease as the device is scaled to future technology nodes. However, to avoid ultrahigh current densities, one must use a thin tunnel barrier that is then subjected to significant voltage-induced stress during repeated writing events [44]. In addition to this increased potential for endurance failure, as the read and write paths are now identical, care must be taken that the read current is large enough to be detectable but small enough to avoid perturbing the state of the tunnel junction [48].

Finally, an exciting alternative to MRAM is to store data in magnetic domain walls by building a magnetic racetrack in the third dimension [49]. Here, domain walls are moved along a loop (using electrical current by means of the spin-torque effect) until they reach an integrated sensor capable of reading and modifying the stored bits [49]. Basic studies are currently underway to understand how domain walls could practically, reliably, and rapidly be moved along such a magnetic racetrack [50], so this technique is probably several years away from being able to evaluate prototype memory cells. The great promise offered by this technique is that by arranging a large number of bits vertically along a deeply etched feature similar to the trench capacitors used in commercial DRAMS, the number of bits per F^2 could potentially be made very large [49].

For more information about MRAM and related spintronics devices, see the issue of the *IBM Journal of Research and Development* on spintronics [45].

Phase-change RAM

Phase-change RAM (PCRAM) exploits the large resistance contrast between the amorphous and crystalline states in phase-change materials [51]. The amorphous phase tends to have high electrical resistivity, while the crystalline phase exhibits a low resistivity, sometimes four or five orders of magnitude lower. Given this large resistance contrast, the difference in read current is more than sufficient for binary storage and even MLC operation [51]. To switch the state of the cell, the phase-change material is crystallized by applying an electrical pulse that heats a significant portion of the cell above its crystallization temperature. This SET operation tends to dictate the write speed performance of PCRAM technology because the required duration of this pulse depends on the crystallization speed of the phase-change material. In the RESET operation, a larger electrical current is applied and then abruptly cut off in order to melt-quench the material, leaving it in the amorphous (high-resistance) state.

Although interest in phase-change memory was slow to develop (Figure 2), a large number of sophisticated integration efforts are now underway in PCRAM technology (see the references that appear in [51]). PCRAM has been shown to offer high endurance [52], fast speed [53], inherent scaling of the phase-change process beyond the 22-nm node [54], and integration at technology nodes down to 90 nm [55]. The most important unknown for the success of PCRAM technology is whether the memory access device (such as

a diode [55] or a transistor [56]) in a dense memory array will be able to supply sufficient current to RESET the phase-change memory cell. Already, in order to try to minimize the RESET current, it is assumed that the dimension of the phase-change material will need to be only 30% of the size of the lithographic dimension, F [2]. However, even with this difficult integration task, the success of PCRAM technology may end up depending on advances in the access device as much as on the phase-change memory cell itself [2].

More details on PCRAM technology and its scaling can be found in the paper by Raoux et al. in this issue [51].

Resistive RAM and solid-electrolyte memory

Over the past 50 years, a large number of materials have been explored for use as a resistive NVM. Although most of these materials can be switched between two distinct resistance states using suitable voltages, the switching mechanism is believed to vary from material to material and is poorly understood. These materials can be classified into two categories on the basis of their operating mechanism—insulator resistive memories and solid-electrolyte (SE) memories. In either case, as with PCRAM, an access device is required to enable the reading and writing of individual memory elements.

Here we chiefly discuss the oxides that are the farthest along in development, namely Cu_xO, NiO, TiO_x, ZrO_x, and HfO_x. The major advantage to using a binary oxide system is the simplicity of the device structure and compatibility with conventional CMOS processing. Important memory attributes of thin Cu_xO have been explored in a 64-Kb memory array at the 180-nm technology node. The switching mechanism is believed to be due to modulation of the space charge limited conduction through occupancy (vacancy) of deep traps [57]. These devices exhibit very fast switching speeds (<50 ns) and low program current (down to 10 μA) but show very poor endurance (600 cycles) and insufficient retention.

Memory cells using the TiO_x material system have been widely studied with a number of different top electrode and bottom electrode materials [58–60]. The application of an electric field to ionic TiO_2 crystals pulls the oxygen ions away from the crystal toward the top electrode, creating oxygen vacancies that form the conductive path in the ON state [58]. For thick TiO_2 (>20 nm) [58], the filament density is extremely low, raising questions about scalability.

Memory cells made with thin TiO_2 (2.5 nm) have also been studied [59] with TiN bottom and Pt top electrodes and have shown fast device switching speeds (<30 ns). However, the RESET current is very high in these devices (~11 mA), which makes a smaller cell size difficult to achieve. In addition, no significant data on endurance or retention exists for this material system. Recently, unipolar operation has been demonstrated in thin TiON using submicron-size cells and switching speeds less than 100 ns [60]. However, the RESET currents are still very high in these devices (more than a few milliamperes) and endurance, retention, and further scalability of the technology are currently unknown.

Nonstoichiometric ZrO_x shows poor device yield [61], and memory cells built with stoichiometric ZrO_2 [62] and nonstoichiometric HfO_x [63] have been hundreds of micrometers wide. In addition, ZrO_2 devices have shown limited endurance and very slow device switching speeds (~1 ms). For the HfO_x system, although low switching voltages and currents have been observed, endurance is very limited (~300 cycles), no high-speed switching data exists, and retention in these devices has not yet been carefully measured. WO_x has also been explored for use as a binary oxide memory [64], and fast switching speeds have been obtained but with very high programming currents.

Early work in the NiO material system demonstrated that the OFF-to-ON transition in this system occurs due to migration of Ni atoms along oxide defects [65] and that the ON-to-OFF transition is due to the thermal rupture of the formed filament. Recent work [66] has also showed scalability down to cell sizes of 0.3 μm × 0.7 μm and a write endurance of 10^6 cycles. However, these devices have only been demonstrated to retain stored data up to 8 months at room temperature and also required a high RESET current (~2 mA), independent of the resistance RAM (RRAM) device area. It has been shown [67] that the RESET current can be reduced to approximately 200 μA in the Pt/NiO/Pt system by limiting both the current that flows during the SET transition and the parasitic currents due to stray capacitance.

Resistive switching phenomena have also been reported for a variety of ternary oxides, including $Pr_{0.7}Ca_{0.3}MnO_3$ (PCMO), (Nb,Cr)-doped $(Ba,Sr)TiO_3$, and $SrZrO_3$, with various top and bottom electrodes [68]. However, for many of these materials, an initial forming process is required before the device exhibits a subsequent well-behaved switching behavior. In addition, there exist significant integration challenges for these more complex materials, and scalability down to less than 100 nm is currently unknown. Cr-doped $SrTiO_3$ is discussed in detail in a separate paper in this issue [69].

In the SE memory system—also referred to as a *programmable metallization cell* (PMC) or the *conductive bridge RAM* [70–72])—an SE material containing mobile metal ions is sandwiched between an inert electrode (cathode) and an oxidizable electrode (anode). A small positive voltage (a few hundred millivolts) at the anode reduces metal ions at the cathode and injects ions into the electrolyte by means of oxidation at the anode. The

electrodeposited filament grows out of the cathode until it contacts the anode, causing the voltage to drop abruptly. A reverse bias of a similar magnitude will erase the device by removing the material by means of reverse electrodeposition, thermal effects, or both. The PMC candidates that are under active investigation include Ag- or Cu-doped germanium chalcogenides (Ge_xSe_{1-x}, Ge_xS_{1-x}, and Ge_xTe_{1-x}) or both; Cu-doped MoO_x (with Cu top electrodes); Cu-doped WO_x; and the $RbAg_4I_5$ system. One of the major advantages of SE memory devices over the various binary resistive memories is the ability to program and erase at very low currents.

The most widely studied PMC system is the Ag-doped Ge_xSe_{1-x} with Ag top electrodes, where Ag is incorporated into the base Ge_xSe_{1-x} glass using ultraviolet diffusion. In this system, scalability down to 20 nm [72], good retention, multilevel capability (made possible by the very high resistance ratio between the ON and OFF states), fast switching speeds (~1 μs or faster), and very high endurance [70] have been successfully demonstrated. However, this system does not survive processing temperatures over 200°C, making integration with a conventional CMOS back end virtually impossible. In contrast, the Ag-doped Ge_xS_{1-x} system has been shown to survive back-end temperatures exceeding 400°C [71], has a high ON/OFF ratio, and also offers fast speeds (~1 μs or faster). Using this system, 2-Mb memory arrays have been fabricated in 90-nm technology with MLC capability [73], but retention and endurance still remain to be demonstrated for these germanium sulfide systems.

Cu-doped Ge_xS_{1-x} devices behave as good SEs as well and promise easier integration with CMOS [71], but show a poorer ON/OFF ratio and have yet to demonstrate tolerance to back-end process temperatures up to 400°C. The Ag-doped Ge_xTe_{1-x} system has also been investigated for PMC devices and shows good temperature stability, but no significant data on endurance or retention exists at this point [74]. Research on Cu-doped MoO_x with Cu top electrodes has shown good endurance (>10^6 cycles) and retention at high program currents for large devices (>100 μm^2), but scalability and lower program currents still remain to be demonstrated [75].

In summary, it appears that the most promising resistive memory candidates for SCM are the Cu_xO and NiO binary metal-oxide systems and the AgGeS SE devices. However, retention and endurance still remain to be demonstrated for the Cu_xO and AgGeS devices, while the high write current is an issue for the NiO memory devices. The AgGeS system has an extremely high ON/OFF ratio and fast switching speeds, but its bipolar program and erase pulses greatly complicate cross-point integration with standard silicon and polysilicon diodes.

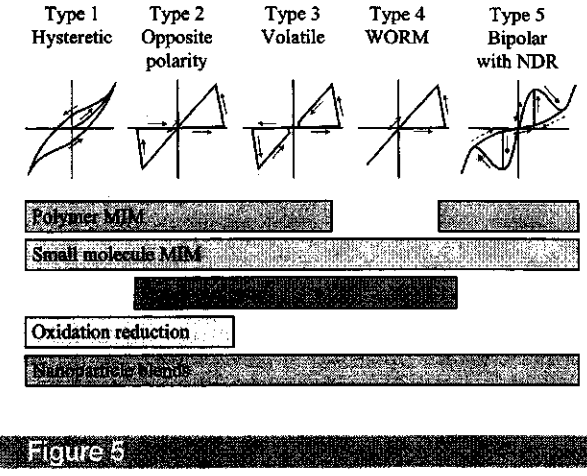

Figure 5

Current–voltage response behavior observed in organic sandwich structures. Only types 2 and 5 are useful as nonvolatile RAM elements.

Organic and polymeric memory

The earliest report of resistive switching behavior in organic two-electrode devices is that of Gregor [76], who used plasma polymerization of divinylbenzene to deposit a film between lead electrodes and observed bistable negative differential resistance (NDR). Despite extensive electrical characterization that included retention time measurement and cycling endurance, it was not possible to identify the mechanism responsible for the switching behavior.

In the four decades since Gregor's studies, a great variety of organic materials has been incorporated into sandwich-like structures that exhibit resistive switching. In a recent review [77], these were categorized in terms of their material components and electrical response (**Figure 5**). In some cases (type 2 in the figure), reversible switching is achieved using voltages of opposite polarity. In systems that show NDR (type 5), the on and off states can be created with voltage pulses of the same polarity but different amplitudes. There are cases in which the conductive state is indefinitely stable (type 4) and would, therefore, serve as a Write-Once, Read-Many (WORM) device. Other responses, though claimed as "memory," either do not retain their state when the voltage is removed (type 3) or are merely hysteretic without the clear threshold that is necessary for incorporation into an addressable array [77].

The simplest memory cell is a two-electrode metal-organic-metal sandwich structure in which the organic layer may be either a polymeric (type 1 in Figure 5) or a low-molecular-weight organic semiconductor (type 2). These can be deposited by solution coating, by thermal evaporation, or by plasma polymerization. The active

Table 2 Candidate technologies for SCM: their technology maturity level, the smallest cell size (highest density) achieved to date, and their prospects for scalability, fast readout, fast writing, low switching power, high endurance, and MLC operation (BJT: bipolar junction transistor).

	Flash	*SONOS flash*	*Nanocrystal flash*	*FeRAM*	*FeFET*
Knowledge level	Product	Advanced development	Development	Product	Basic research
Smallest demonstrated cell	4 F^2 (2 F^2 per bit [2])	4 F^2 (1 F^2 per bit [21])	16 F^2 at 90 nm [2, 24]	15 F^2 at 130 nm [40]	—
Prospects for:					
Scalability	Poor	Maybe (enough stored charge?)	Unclear (enough stored charge?)	Poor (integration, signal loss)	Unclear (difficult integration)
Fast readout	Yes	Yes	Yes	Yes	Yes
Fast writing	No	No	No	Yes	Yes
Low switching power	Yes	Yes	Yes	Yes	Yes
High endurance	No	Poor (10^7 cycles [2])	No	Yes	Yes
Nonvolatility	Yes	Yes	Yes	Yes	Poor (30 days [2])
MLC operation	Yes	Yes (4 bits [21])	Yes	Difficult	Difficult

layer is assumed to be initially homogeneous and is typically highly resistive in the pristine device, hence a metal-insulator-metal, or *MIM*. In type 3, the active layer consists of an electron donor, such as carbazoles [78] or arylamines [79], together with an acceptor, such as TCNQ [80] or a fullerene [81]. Depending on the oxidation potential of the donor and the reduction potential of the acceptor, in the ground state the complex may have both species nominally neutral or fully charge-transferred, with both cases resulting in high resistance. Partial charge transfer can then lead to a more highly conductive state. Closely related are type 4 electrochemical systems in which oxidation and reduction at the electrodes, accompanied by ion migration [82], change the conductivity of the organic material.

A great deal of recent study has been devoted to memory elements consisting of metallic nanoparticles (NPs) blended into an organic semiconducting host (type 5). The NPs may be in a relatively discrete layer [83] or mixed more uniformly throughout the organic [84].

Just as in inorganic systems, many kinds of physical and chemical changes have been invoked—but rarely proved—for the creation of the conducting state: formation of metallic filaments due to electromigration from an electrode [85]; carbon filaments due to pyrolysis of the organic material in electrical breakdown [86]; realignment of molecular species to permit higher mobility, for example, by pi-pi stacking [87]; or electrochemical oxidation or reduction. Modification of the electrode–organic interface with a corresponding change in charge injection has also been proposed as the mechanism for switching [88].

Many NP-blend devices exhibit bipolar switching with NDR above a sharp threshold (type 5 in Figure 5). This is very similar to the characteristics of current and voltage interaction observed by Simmons and Verderber [89] in inorganic MIMs and attributed to charging effects in gold clusters. The NPs act as traps that can be charged and discharged by suitable voltage pulses. Charge retention [90] of at least weeks has been demonstrated. The details of this mechanism remain elusive: Tang et al. [91] showed that accidental NPs may be present in nominally homogeneous organic layers, thereby accounting for the similarity in response of apparently disparate devices, and suggested that particular NP configurations [92] are responsible for switching.

Optimizing the characteristics that are desirable for memory applications first requires clarification of the switching mechanism. NP blends show promising data-retention times, switching speed, and cycling endurance, but the on-state current is too low to permit scaling to nanometer dimensions. In addition, the lack of significant rectification (they make poor diodes) prevents direct use in a cross-point array.

In contrast to other NVM candidates, there have not been many demonstrations of integrated device arrays or device-centric demonstrations (Figure 2). This makes it more difficult to assess the possible performance of organic memory devices in SCM. Certainly, such demonstrations would help indicate whether these materials could be integrated in a conventional CMOS process (even to CMOS back-end temperatures of perhaps 350–400°C) and the prospects of device scalability.

Table 2 Continued.

MRAM	*Racetrack*	*PCRAM*	*RRAM*	*Solid electrolyte*	*Organic memory*
Product	Basic research	Advanced development	Early development	Development	Basic research
25 F^2 at 180 nm [2, 45]	—	5.8 F^2 diode [55] 12 F^2 BJT [56] at 90 nm	—	8 F^2 at 90 nm [73] (4 F^2 per bit)	—
Poor (high currents)	Unknown (high current density?)	Promising (rapid progress to date)	Unknown	Promising (filament based, but must integrate new materials)	Unknown (back-end temperatures?)
Yes	Yes	Yes	Yes	Yes	Sometimes
Yes	Yes	Yes	Sometimes	Yes	Sometimes
No	Uncertain	Poor	Sometimes	Yes	Sometimes
Yes	Unknown	Yes	Poor	Unknown	Poor
Yes	Unknown	Yes	Sometimes	Sometimes	Poor
No	Yes (3D)	Yes	Yes	Yes	Unknown

Summary and outlook

We have discussed a number of candidates currently under investigation as possible NVM technologies, including flash alternatives such as SONOS and nanocrystal flash, FeRAM, MRAM, PCRAM, RRAM, SEs, and organic and polymer memory. In **Table 2**, we qualitatively assess these candidates in terms of the aggressive target specifications laid out in Table 1.

The current knowledge level for each candidate NVM technology ranges from well-known technologies being shipped as product, to development efforts designed to implement integrated cells and optimize device parameters, to basic research efforts into underlying mechanisms. If integrated test arrays are known to have been produced, then the smallest cell size achieved is indicated in the second row of Table 2.

In the remainder of Table 2, we briefly summarize the prospects of each NVM technology candidate for achieving the scalability, fast readout, fast writing, low switching power, high endurance, and MLC operation that would be necessary in order to deliver a successful SCM. Although many of these aspects are often closely interrelated, this table serves as a quick guide to which aspects could become insurmountable obstacles for each candidate SCM technology.

For instance, while SONOS flash and nanocrystal flash are clearly designed to improve the scalability of flash to further technology nodes, neither of them is expected to greatly improve the other two weaknesses of flash technology, slow writing and relatively poor endurance. In contrast, FeRAM and MRAM demonstrate high speed and endurance, but their inherent scaling limitations have not been overcome. In fact, both have spawned alternative approaches (FeFETs and racetrack memory, respectively) expressly to avoid these difficulties. However, each of these new techniques is sufficiently different that more research will be required to fully assess their prospects. In one sense, racetrack and spin-torque memories would seem to be more promising partly because FeFETs are an old concept that has long been known to be difficult [37–38] and partly because racetrack memory [49] offers the built-in prospect of storing many bits per F^2 by harnessing a recently discovered physical phenomenon [47].

Since RRAM and organic memory offer so many variants with widely different observed behaviors, for each desirable performance aspect, one can often find an example that satisfies the criteria and another that fails. However, it is generally true that no one variant exhibits characteristics that are all favorable, and in general, ultrahigh endurance has not yet been convincingly demonstrated in either RRAM or organic memory.

This leaves PCRAM and SEs as the current frontrunners for SCM. In both cases, the scaling issues appear to primarily involve expected difficulties with incorporating novel materials into a CMOS fabrication process. Both SEs and phase-change have been shown to demonstrate basic functionality at extremely small feature sizes: in 20-nm cylindrical SE vias [72] and in phase-change bridges of 3 nm × 20 nm in cross-section [54]. This is shown by the open symbols on **Figure 6**, which are compared to the trends for flash scaling in terms of effective area per bit as a function of year [2]. Here we have assumed that an integrated cell incorporating these

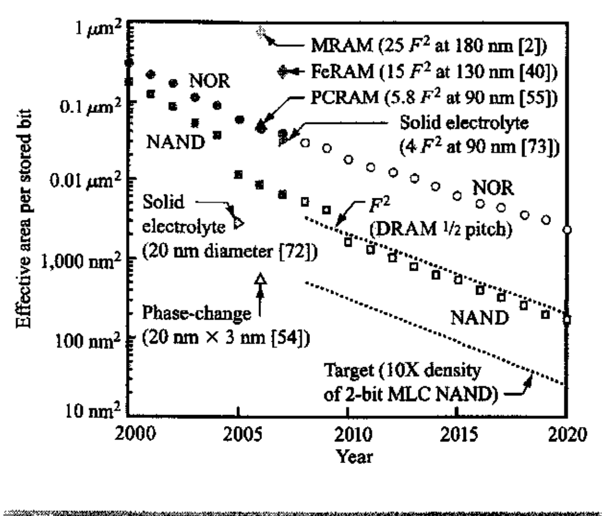

Figure 6

Effective area per stored bit for NAND and NOR flash as a function of year. Also shown are filled symbols corresponding to some recent demonstrations of fully integrated memory arrays for various SCM candidate technologies and open symbols corresponding to some single-device demonstrations illustrating raw scaling potential.

ultrascaled areas would likely be three times larger in diameter than the device aperture itself. Also shown, as filled symbols, are several recent demonstrations of integrated test arrays. We note that two jumps are indicated in NAND best-case effective area per bit, one for the advent of 2-bit MLC, and a second for the projected move to 4-bit MLC [2]. At that point, the effective bit size for NAND flash will have scaled faster than lithography, as indicated by the dotted line for DRAM half-pitch.

In PCRAM, the high current required to RESET the cell will have to match the small current-sourcing capabilities of a scaled-down access device (such as a diode or transistor). SE materials must add tolerance to back-end processing temperatures (350–400°C) to the otherwise favorable aspects of earlier variants (Ag-GeSe) and improve the data retention of the thin filaments formed by low programming currents without relying on read-induced reinforcement.

We note that with the promising scaling behavior of both phase-change and SE memory, recent integrated device demonstrations have essentially caught up with NOR flash despite being implemented at 90-nm technology (instead of at 65 nm). However, the target density for SCM, indicated on Figure 6 with a red dotted line, will require densities nearly two orders of magnitude higher. The red line assumes a 10-fold price difference between 2-bit MLC NAND flash and high-performance server HDDs, given the curves for the lower-cost desktop HDDs shown in Figure 1.

However, even partway toward meeting the target SCM specifications (Table 1), such an NVM would already be attractive to existing market segments. If the test vehicles being built with phase-change and SE memory (already competitive with NOR flash in terms of density) can be transferred to high-yield manufacturing of memory chips and satisfy the stringent error-rate requirements of NOR flash, then these emerging NVM technologies could compete with NOR flash in the near future. In turn, as the increasing density of such a new NVM technology drives down cost, markets for NAND flash, DRAM (if speeds could be made high enough), and server-class HDD storage could potentially be addressed. Along this path, the increasing size of these markets would serve to motivate investment of the resources required for the next round of technology development. If the full promise of SCM can be realized, we could witness the birth of the first truly universal memory, capable of supplanting everything in the memory and storage hierarchy between L1 cache DRAM and magnetic tape.

**Trademark, service mark, or registered trademark of Sony Computer Entertainment Corporation in the United States, other countries, or both.

References

1. G. E. Moore, "Cramming More Components onto Integrated Circuits," *Electronics* **38**, No. 8, 114–117 (1965).
2. International Technology Roadmap for Semiconductors, ITRS 2006 Update; see *http://www.itrs.net/Links/2006Update/2006UpdateFinal.htm*.
3. K. Kim and J. Choi, "Future Outlook of NAND Flash Technology for 40nm Node and Beyond," *Proceedings of the IEEE Non-Volatile Semiconductor Memory Workshop*, Monterey, CA, 2006, pp. 9–11.
4. S. K. Lai, "Flash Memories: Successes and Challenges," *IBM J. Res. & Dev.* **52**, No. 4/5, 529–535 (2008, this issue).
5. R. H. Dennard, F. H. Gaensslen, V. L. Rideout, E. Bassous, and A. R. LeBlanc, "Design of Ion-Implanted MOSFET's with Very Small Physical Dimensions," *IEEE J. Solid-State Circ.* **9**, No. 5, 256–268 (1974).
6. E. Grochowski and R. D. Halem, "Technological Impact of Magnetic Hard Disk Drives on Storage Systems," *IBM Syst. J.* **42**, No. 2, 338–346 (2003).
7. D. A. Patterson, "Latency Lags Bandwidth," *Comm. ACM* **47**, No. 10, 71–75 (2004).
8. R. F. Freitas and W. W. Wilcke, "Storage-Class Memory: The Next Storage System Technology," *IBM J. Res. & Dev.* **52**, No. 4/5, 439–447 (2008, this issue).
9. M. Johnson, A. Al-Shamma, D. Bosch, M. Crowley, M. Farmwald, L. Fasoli, A. Ilkbahar, et al., "512-Mb PROM with a Three-Dimensional Array of Diode/Antifuse Memory Cells," *IEEE J. Solid-State Circ.* **38**, No. 11, 1920–1928 (2003).
10. B. Eitan, R. Kazerounian, A. Roy, G. Crisenza, P. Cappelletti, and A. Modelli, "Multilevel Flash Cells and Their Trade-offs," *Proceedings of the IEEE International Electron Devices Meeting*, San Francisco, CA, 1996, pp. 169–172.
11. K. Gopalakrishnan, R. S. Shenoy, C. T. Rettner, R. S. King, Y. Zhang, B. Kurdi, L. D. Bozano, et al., "The Micro to Nano Addressing Block (MNAB)," *Proceedings of the IEEE*

International Electron Devices Meeting, Washington, DC, 2005, pp. 471–474.

12. G. Campardo and R. Micheloni, "Special Issue on Flash Technology," *Proc. IEEE* **91**, No. 4, 483–488 (2003).
13. Toshiba America Electronic Components, Inc., "NAND vs. NOR Flash Memory: Technology Overview"; see *http://www.toshiba.com/taec/components/Generic/Memory_Resources/NANDvsNOR.pdf.*
14. Micron Technology, Inc., "NAND Flash 101: An Introduction to NAND Flash and How to Design It in to Your Next Product," Technical Note No. TN-29-19; see *http://download.micron.com/pdf/technotes/nand/tn2919.pdf.*
15. iSuppli Corporation; see *http://isuppli.com.*
16. L. Mason, "Memory Market Outlook: DRAMs Are Nice, But All the Action Is in NAND Flash," *Proceedings of the Denali MemCon Conference*, Shanghai, China, 2007; see *http://www.bloobble.com/objects/presentations?itemid=454.*
17. K. Kim, "Technology for Sub-50nm DRAM and NAND Flash Manufacturing," *Proceedings of the IEEE International Electron Devices Meeting*, Washington, DC, 2005, pp. 323–326.
18. J.-D. Lee, S.-H. Hur, and J.-D. Choi, "Effects of Floating-Gate Interference on NAND Flash Memory Cell Operation," *IEEE Elect. Dev. Lett.* **23**, No. 5, 264–266 (2002).
19. M. H. White, D. A. Adams, and J. Bu, "On the Go with SONOS," *IEEE Circ. & Dev.* **16**, No. 4, 22–31 (2000).
20. W. J. Tsai, N. K. Zous, C. J. Liu, C. C. Liu, C. H. Chen, W. Tahui, S. Pan, C.-Y. Lu, and S. H. Gu, "Data Retention Behavior of a SONOS Type Two-Bit Storage Flash Memory Cell," *Proceedings of the IEEE International Electron Devices Meeting*, Washington, DC, 2001, pp. 32.6.1–32.6.4.
21. C. W. Oh, N. Y. Kim, S. H. Kim, Y. L. Choi, S. I. Hong, H. J. Bae, J. B. Kim, et al., "4-Bit Double SONOS Memories (DSMs) Using Single-Level and Multi-Level Cell Schemes," *Proceedings of the IEEE International Electron Devices Meeting*, San Francisco, CA, 2006, pp. 1–4.
22. C. H. Lee, K. I. Choi, M. K. Choi, Y. H. Song, K. C. Park, and K. Kim, "A Novel SONOS Structure of $SiO_2/SiN/Al_2O_3$ with TaN Metal Gate for Multi-giga Bit Flash Memories," *Proceedings of the IEEE International Electron Devices Meeting*, Washington, DC, 2003, pp. 26.5.1–26.5.4.
23. Z. L. Huo, J. K. Yang, S. H. Lim, S. J. Baik, J. Lee, J. H. Han, I.-S. Yeo, U.-I. Chung, J. T. Moon, and B.-I. Ryu, "Band Engineered Charge Trap Layer for Highly Reliable MLC Flash Memory," *Proceedings of the IEEE Symposium on VLSI Technology*, Kyoto, Japan, 2007, pp. 138–139.
24. R. Muralidhar, R. F. Steimle, M. Sadd, R. Rao, C. T. Swift, E. J. Prinz, J. Yater, et al., "A 6 V Embedded 90 nm Silicon Nanocrystal Nonvolatile Memory," *Proceedings of the IEEE International Electron Devices Meeting*, Washington, DC, 2003, pp. 26.2.1–26.2.4.
25. B. Hwang, J. Shim, J.-H. Park, K. Lee, S. Kwon, S.-Y. Park, Y. Park, et al., "Development of Bit-Line Contact of 76 nm Pitch on NAND Flash Cell Using Reversal PR (Photo Resist) and SADP (Self-Align Double Patterning) Process," *Proceedings of the 18th Annual IEEE/SEMI Advanced Semiconductor Manufacturing Conference*, Stresa, Italy, 2007, pp. 356–358 (2007).
26. T. Endoh, K. Kinoshita, T. Tanigami, Y. Wada, K. Sato, K. Yamada, T. Yokoyama, et al., "Novel Ultrahigh-Density Flash Memory with a Stacked-Surrounding Gate Transistor (S-SGT) Structured Cell," *IEEE Trans. Elect. Dev.* **50**, No. 4, 945–951 (2003).
27. H. Tanaka, M. Kido, K. Yahashi, M. Oomura, R. Katsumata, M. Kito, Y. Fukuzumi, et al., "Bit Cost Scalable Technology with Punch and Plug Process for Ultra High Density Flash Memory," *Proceedings of the IEEE Symposium on VLSI Technology*, Kyoto, Japan, 2007, pp. 14–15.
28. P. Xuan, M. She, B. Harteneck, A. Liddle, J. Bokor, and T.-J. King, "FinFET SONOS Flash Memory for Embedded Applications," *Proceedings of the IEEE International Electron Devices Meeting*, Washington, DC, 2003, pp. 26.4.1–26.4.4.
29. A. Sheikholeslami and P. G. Gulak, "A Survey of Circuit Innovations in Ferroelectric Random-access Memories," *Proc. IEEE* **88**, No. 5, 667–689 (2000).
30. Y. Kato, T. Yamada, and Y. Shimada, "0.18-μm Nondestructive Readout FeRAM Using Charge Compensation Technique," *IEEE Trans. Elect. Dev.* **52**, No. 12, 2616–2621 (2005).
31. K. Kim and S. Lee, "Integration of Lead Zirconium Titanate Thin Films for High Density Ferroelectric Random Access Memory," *J. Appl. Phys.* **100**, No. 5, 051604 (2006).
32. M. Dawber, K. M. Rabe, and J. F. Scott, "Physics of Thin-Film Ferroelectric Oxides," *Rev. Mod. Phys.* **77**, No. 4, 1083–1130 (2005).
33. N. Setter, D. Damjanovic, L. Eng, G. Fox, S. Gevorgian, S. Hong, A. Kingon, et al., "Ferroelectric Thin Films: Review of Materials, Properties, and Applications," *J. Appl. Phys.* **100**, No. 5, 051606–051646 (2006).
34. K. Ashikaga, K. Takaya, T. Kanehara, M. Yoshimaru, and I. Koiwa, "Reduction of Process-induced Damage and Improvement of Imprint Characteristics in $SrBi_2Ta_2O_9$ Capacitors by Post Metallization Annealing," *Japan. J. Appl. Phys.* **46**, No. 2, 695–697 (2007).
35. H. J. Joo, S. K. Kang, J. H. Park, H. S. Kim, J. H. Kim, J. Y. Jung, D. Y. Choi, et al., "A Novel ATE (Additional Top-Electrode) Scheme for a 1.6 V FRAM Embedded Device at 180 nm Technology," *Int. Ferro.* **89**, No. 1, 106–115 (2007).
36. D. Takashima and I. Kunishima, "High-Density Chain Ferroelectric Random Access Memory (Chain FRAM)," *IEEE J. Solid-State Circ.* **33**, No. 5, 787–792 (1998).
37. S.-Y. Wu, "A New Ferroelectric Memory Device, Metal-Ferroelectric-Semiconductor Transistor," *IEEE Trans. Elect. Dev.* **21**, No. 8, 499–504 (1974).
38. T. P. Ma and J.-P. Han, "Why Is Nonvolatile Ferroelectric Memory Field-Effect Transistor Still Elusive?" *IEEE Elect. Dev. Lett.* **23**, No. 7, 386–388 (2002).
39. N. Tanabe, S. Kobayashi, T. Miwa, K. Amamuma, H. Mori, N. Inoue, T. Takeuchi, et al., "High Tolerance Operation of 1T/2C FeRAMs for the Variation of Cell Capacitors Characteristics," *Proceedings of the IEEE Symposium on VLSI Technology*, Honolulu, HI, 1998, pp. 124–125.
40. Y. K. Hong, D. J. Jung, S. K. Kang, H. S. Kim, J. Y. Jung, H. K. Koh, J. H. Park, et al., "130 nm-Technology, 0.25 μm^2, 1T1C FRAM Cell for SoC (System-on-a-Chip)-friendly Applications," *Proceedings of the IEEE Symposium on VLSI Technology*, Kyoto, Japan, 2007, pp. 230–231.
41. K. R. Udayakumar, T. S. Moise, S. R. Summerfelt, K. Boku, K. A. Remack, J. Gertas, A. Haider, et al., "Full-Bit Functional, High-Density 8 Mbit One Transistor-One Capacitor Ferroelectric Random Access Memory Embedded within a Low-power 130 nm Logic Process," *Japan. J. Appl. Phys.* **46**, No. 4B, 2180–2183 (2007).
42. S. Parkin, X. Jiang, C. Kaiser, A. Panchula, K. Roche, and M. Samant, "Magnetically Engineered Spintronic Sensors and Memory," *Proc. IEEE* **91**, No. 5, 661–680 (2003).
43. M. Julliere, "Tunneling Between Ferromagnetic-Films," *Phys. Lett. A* **54**, No. 3, 225–226 (1975).
44. M. Durlam, Y. Chung, M. DeHerrera, B. N. Engel, G. Grynkewich, B. Martino, B. Nguyen, J. Salter, P. Shah, and J. M. Slaughter, "MRAM Memory for Embedded and Stand Alone Systems," *Proceedings of the IEEE International Conference on Integrated Circuit Design and Technology*, Austin, TX, 2007, pp. 1–4.
45. "Spintronics," *IBM J. Res. & Dev.* **50**, No. 1 (2006, entire issue).
46. I. L. Prejbeanu, W. Kula, K. Ounadjela, R. C. Sousa, O. Redon, B. Dieny, and J.-P. Nozieres, "Thermally Assisted Switching in Exchange-Biased Storage Layer Magnetic Tunnel Junctions," *IEEE Trans. Magn.* **40**, No. 4, 2625–2627 (2004).
47. J. C. Slonczewski, "Current-Driven Excitation of Magnetic Multilayers," *J. Magn. Magn. Mat.* **159**, No. 1/2, L1–L7 (1996).

48. K. Miura, T. Kawahara, R. Takemura, J. Hayakawa, S. Ikeda, R. Sasaki, H. Takahashi, H. Matsuoka, and H. Ohno, "A Novel SPRAM (SPin-Transfer Torque RAM) with a Synthetic Ferrimagnetic Free Layer for Higher Immunity to Read Disturbance and Reducing Write-Current Dispersion," *Proceedings of the IEEE Symposium on VLSI Technology*, Kyoto, Japan, pp. 234–235 (2007).
49. S. S. P. Parkin, "Spintronic Materials and Devices: Past, Present and Future!" *Proceedings of the IEEE International Electron Devices Meeting*, San Francisco, CA, 2004, pp. 903–906.
50. L. Thomas, M. Hayashi, X. Jiang, R. Moriya, C. Rettner, and S. Parkin, "Resonant Amplification of Magnetic Domain-Wall Motion by a Train of Current Pulses," *Science* **315**, No. 5818, 1553–1556 (2007).
51. S. Raoux, G. W. Burr, M. J. Breitwisch, C. T. Rettner, Y.-C. Chen, R. M. Shelby, M. Salinga, et al., "Phase-Change Random Access Memory: A Scalable Technology," *IBM J. Res. & Dev.* **52**, No. 4/5, 465–479 (2008, this issue).
52. S. Lai, "Current Status of the Phase Change Memory and its Future," *Proceedings of the IEEE International Electron Devices Meeting*, Washington, D.C., 2003, pp. 10.1.1–10.1.4.
53. A. Pirovano, A. L. Lacaita, A. Benvenuti, F. Pellizzer, S. Hudgens, and R. Bez, "Scaling Analysis of Phase-Change Memory Technology," *Proceedings of the IEEE International Electron Devices Meeting*, Washington, DC, 2003, pp. 29.6.1–29.6.4.
54. Y. C. Chen, C. T. Rettner, S. Raoux, G. W. Burr, S. H. Chen, R. M. Shelby, M. Salinga, et al., "Ultra-thin Phase-change Bridge Memory Device Using GeSb," *Proceedings of the IEEE International Electron Devices Meeting*, San Francisco, CA, 2006, pp. 30.3.1–30.3.4.
55. J. H. Oh, J. H. Park, Y. S. Lim, H. S. Lim, Y. T. Oh, J. S. Kim, J. M. Shin, et al., "Full Integration of Highly Manufacturable 512Mb PRAM Based on 90nm Technology," *Proceedings of the IEEE International Electron Devices Meeting*, San Francisco, CA, 2006, pp. 2.6.1–2.6.4.
56. F. Pellizzer, A. Benvenuti, B. Gleixner, Y. Kim, B. Johnson, M. Magistretti, T. Marangon, A. Pirovano, R. Bez, and G. Atwood, "A 90nm Phase Change Memory Technology for Stand-alone Non-volatile Memory Applications," *Proceedings of the IEEE Symposium on VLSI Technology*, Honolulu, HI, 2006, pp. 122–123.
57. A. Chen, S. Haddad, Y.-C. Wu, T.-N. Fang, Z. Lan, S. Avanzino, S. Pangrle, et al., "Non-volatile Resistive Switching for Advanced Memory Applications," *Proceedings of the IEEE International Electron Devices Meeting*, Washington, DC, 2005, pp. 746–749.
58. B. J. Choi, D. S. Jeong, S. K. Kim, C. Rohde, S. Choi, J. H. Oh, H. J. Kim, et al., "Resistive Switching Mechanism of TiO_2 Thin Films Grown by Atomic-Layer Deposition," *J. Appl. Phys.* **98**, No. 3, 033715 (2005).
59. M. Fujimoto, H. Koyama, Y. Hosoi, K. Ishihara, and S. Kobayashi, "High-Speed Resistive Switching of TiO_2/TiN Nano-crystalline Thin Film," *Japan. J. Appl. Phys.* **45**, No. 8/11, L310–L312 (2006).
60. Y. Hosoi, Y. Tamai, T. Ohnishi, K. Ishihara, T. Shibuya, Y. Inoue, S. Yamazaki, et al., "High Speed Unipolar Switching Resistance RAM (RRAM) Technology," *Proceedings of the IEEE International Electron Devices Meeting*, San Francisco, CA, 2006, pp. 30.7.1–30.7.4.
61. D. Lee, D.-J. Seong, H. J. Choi, I. Jo, R. Dong, W. Xiang, S. Oh, et al., "Excellent Uniformity and Reproducible Resistance Switching Characteristics of Doped Binary Metal Oxides for Non-volatile Resistance Memory Applications," *Proceedings of the IEEE International Electron Devices Meeting*, San Francisco, CA, 2006, pp. 30.8.1–30.8.4.
62. X. Wu, P. Zhou, J. Li, L. Y. Chen, H. B. Lin, Y. Y. Lin, and T. A. Tang, "Reproducible Unipolar Resistance Switching in Stoichiometric ZrO_2 Films," *Appl. Phys. Lett.* **90**, No. 18, 183507–183510 (2007).
63. H.-Y. Lee, P.-S. Chen, C.-C. Wang, S. Maikap, P.-J. Tzeng, C.-H. Lin, L.-S. Lee, and M.-J. Tsai, "Low-Power Switching of Nonvolatile Resistive Memory Using Hafnium Oxide," *Japan. J. Appl. Phys.* **46**, No. 4B, 2175–2179 (2007).
64. C. H. Ho, E. K. Lai, M. D. Lee, C. L. Pan, Y. D. Yao, K. Y. Hsieh, R. Liu, and C. Y. Lu, "A Highly Reliable Self-aligned Graded Oxide WO_x Resistance Memory: Conduction Mechanisms and Reliability," *Proceedings of the IEEE Symposium on VLSI Technology*, Kyoto, Japan, 2007, pp. 228–229.
65. J. F. Gibbons and W. E. Beadle, "Switching Properties of Thin NiO Films," *Solid State Elect.* **7**, No. 11, 785–790 (1964).
66. I. G. Baek, M. S. Lee, S. Seo, M. J. Lee, D. H. Seo, D.-S. Suh, J. C. Park, et al., "Highly Scalable Nonvolatile Resistive Memory Using Simple Binary Oxide Driven by Asymmetric Unipolar Voltage Pulses," *Proceedings of the IEEE International Electron Devices Meeting*, San Francisco, CA, 2004, pp. 587–590.
67. K. Kinoshita, K. Tsunoda, Y. Sato, H. Noshiro, Y. Yamazaki, T. Fukano, S. Yagaki, M. Aoki, and Y. Sugiyama, "Reduction of Reset Current in NiO-ReRAM Brought About by Ideal Current Limiter," *Proceedings of the 22nd IEEE Non-Volatile Semiconductor Memory Workshop*, Monterey, CA, 2007, pp. 66–67.
68. A. Beck, J. G. Bednorz, C. Gerber, C. Rossel, and D. Widmer, "Reproducible Switching Effect in Thin Oxide Films for Memory Applications," *Appl. Phys. Lett.* **77**, No. 1, 139–141 (2000).
69. S. F. Karg, G. I. Meijer, J. G. Bednorz, C. T. Rettner, A. G. Schrott, E. A. Joseph, C. H. Lam, et al., "Transition-Metal-Oxide-Based Resistance-Change Memories," *IBM J. Res. & Dev.* **52**, No. 4/5, 481–492 (2008, this issue).
70. M. N. Kozicki, M. Park, and M. Mitkova, "Nanoscale Memory Elements Based on Solid-state Electrolytes," *IEEE Trans. Nano.* **4**, No. 3, 331–338 (2005).
71. M. N. Kozicki, M. Balakrishnan, C. Gopalan, C. Ratnakumar, and M. Mitkova, "Programmable Metallization Cell Memory Based on Ag-Ge-S and Cu-Ge-S Solid Electrolytes," *Proceedings of the 20th IEEE Non-Volatile Semiconductor Memory Workshop*, Monterey, CA, 2005, pp. 83–89.
72. M. Kund, G. Beitel, C.-U. Pinnow, T. Röhr, J. Schumann, R. Symanczyk, K.-D. Ufert, and G. Müller, "Conductive Bridging RAM (CBRAM): An Emerging Non-volatile Memory Technology Scalable to Sub 20nm," *Proceedings of the IEEE International Electron Devices Meeting*, Washington, DC, 2005, pp. 754–757 (2005).
73. P. Schrögmeier, M. Angerbauer, S. Dietrich, M. Ivanov, H. Hönigschmid, C. Liaw, M. Markert, et al., "Time Discrete Voltage Sensing and Iterative Programming Control for a $4F^2$ Multilevel CBRAM," *Proceedings of the IEEE Symposium on VLSI Circuits*, Kyoto, Japan, pp. 186–187 (2007).
74. C. J. Kim, S. G. Yoon, K. J. Choi, S. O. Ryu, S. M. Yoon, N. Y. Lee, and B. G. Yu, "Characterization of Silver-saturated Ge-Te Chalcogenide Thin Films for Nonvolatile Random Access Memory," *J. Vac. Sci. & Tech. B* **24**, No. 2, 721–724 (2006).
75. D. Lee, D.-J. Seong, I. Jo, F. Xiang, R. Dong, S. Oh, and H. Hwang, "Resistance Switching of Copper Doped MoO_x Films for Nonvolatile Memory Applications," *Appl. Phys. Lett.* **90**, No. 12, 122104 (2007).
76. L. V. Gregor, "Polymer Dielectric Films," *IBM J. Res. & Dev.* **12**, No. 2, 140–162 (1968).
77. J. C. Scott and L. D. Bozano, "Nonvolatile Memory Elements Based on Organic Materials," *Adv. Mater.* **19**, 1452–1463 (2007).
78. Q. Ling, Y. Song, S. J. Ding, C. Zhu, D. S. H. Chan, D.-L. Kwong, E.-T. Kang, and K.-G. Neoh, "Non-volatile Polymer Memory Device Based on a Novel Copolymer of *N*-vinylcarbazole and Eu-Complexed Vinylbenzoate," *Adv. Mater.* **17**, No. 4, 455–459 (2005).

79. B. Pradhan, S. K. Batabyal, and A. J. Pal, "Electrical Bistability and Memory Phenomenon in Carbon Nanotube-conjugated Polymer Matrixes," *J. Phys. Chem. B* **110**, No. 16, 8274–8277 (2006).
80. R. S. Potember, T. O. Poehler, and D. O. Cowan, "Electrical Switching and Memory Phenomena in Cu-TCNQ Thin Films," *Appl. Phys. Lett.* **34**, No. 6, 405–407 (1979).
81. C.-W. Chu, J. Ouyang, J.-H. Tseng, and Y. Yang, "Organic Donor-Acceptor System Exhibiting Electrical Bistability for Use in Memory Devices," *Adv. Mater.* **17**, 1440–1443 (2005).
82. F. Verbakel, S. C. J. Meskers, and R. A. J. Janssen, "Electronic Memory Effects in a Sexithiophene-poly(ethylene oxide) Block Copolymer Doped with NaCl Combined Diode and Resistive Switching Behavior," *Chem. Mater.* **18**, No. 11, 2707–2712 (2006).
83. L. D. Bozano, B. W. Kean, V. R. Deline, J. R. Salem, and J. C. Scott, "Mechanism for Bistability in Organic Memory Elements," *Appl. Phys. Lett.* **84**, No. 4, 607–609 (2004).
84. L. D. Bozano, B. W. Kean, M. Beinhoff, K. R. Carter, P. M. Rice, and J. C. Scott, "Organic Materials and Thin-Film Structures for Cross-point Memory Cells Based on Trapping in Metallic Nanoparticles," *Adv. Func. Mater.* **15**, No. 12, 1933–1939 (2005).
85. G. Dearnaley, D. V. Morgan, and A. M. Stoneham, "A Model for Filament Growth and Switching in Amorphous Oxide Films," *J. Non-Cryst. Solids* **4**, 593–612 (1970).
86. H. Pagnia and N. Sotnik, "Bistable Switching in Electroformed Metal-Insulator-Metal Devices," *Phys. Stat. Sol. A* **108**, No. 1, 11–65 (1988).
87. A. Bandyopadhyay and A. J. Pal, "Key to Design Functional Organic Molecules for Binary Operation with Large Conductance Switching," *Chem. Phys. Lett.* **371**, No. 1/2, 86–90 (2003).
88. D. M. Taylor and C. A. Mills, "Memory Effect in the Current–Voltage Characteristic of a Low-Band Gap Conjugated Polymer," *J. Appl. Phys.* **90**, No. 1, 306–309 (2001).
89. J. G. Simmons and R. R. Verderber, "New Conduction and Reversible Memory Phenomena in Thin Insulating Films," *Proc. Royal Soc. London A* **301**, No. 1464, 77–102 (1967).
90. Y. Yang, J. Ouyang, L. Ma, R. J.-H. Tseng, and C.-W. Chu, "Electrical Switching and Bistability in Organic/Polymeric Thin Films and Memory Devices," *Adv. Funct. Mater.* **16**, 1001–1014 (2006).
91. W. Tang, H. Z. Shi, G. Xu, B. S. Ong, Z. D. Popovic, J. C. Deng, J. Zhao, and G. H. Rao, "Memory Effect and Negative Differential Resistance by Electrode-Induced Two-dimensional Single-Electron Tunneling in Molecular and Organic Electronic Devices," *Adv. Mater.* **17**, No. 19, 2307–2311 (2005).
92. M. Shin, S. Lee, K. W. Park, and E.-H. Lee, "Secondary Coulomb Blockade Gap in a Four-Island Tunnel-Junction Array," *Phys. Rev.* **59**, No. 4, 3160–3167 (1999).

Received October 1, 2007; accepted for publication December 31, 2007

Geoffrey W. Burr *IBM Almaden Research Center, 650 Harry Road, San Jose, California 95120 (burr@almaden.ibm.com).* Dr. Burr received his B.S. degree in electrical engineering from the State University of New York at Buffalo in 1991 and his M.S. and Ph.D. degrees in electrical engineering from the California Institute of Technology in 1993 and 1996, respectively. Since that time, Dr. Burr has worked at the IBM Almaden Research Center, where he is currently a Research Staff Member. After having worked for a number of years as an experimentalist in volume holographic data storage and optical information processing, his current research interests include nanophotonics, numerical modeling for design optimization, phase-change memory, and other nonvolatile memory.

Bülent N. Kurdi *IBM Almaden Research Center, 650 Harry Road, San Jose, California 95120 (bulent@us.ibm.com).* Dr. Kurdi completed his Ph.D. studies at the Institute of Optics at the University of Rochester where he investigated silicon-based integrated optics. He holds B.S. degrees in electrical engineering and mathematics with minors in physics and philosophy from the University of Dayton. He joined the IBM Almaden Research Center in 1989 where he has worked on integrated optical devices for magneto-optical data storage, top surface imaging techniques for the fabrication of advanced magnetic write heads, and planarization processes for magnetic head slider fabrication. He is currently the manager of the nanoscale device integration group and has been coordinating several multifaceted efforts in the area of ultrahigh-density NVM devices.

J. Campbell Scott *IBM Almaden Research Center, 650 Harry Road, San Jose, California 95120 (jcscott@almaden.ibm.com).* Dr. Scott is a Research Staff Member. He received a B.S. degree in physics from St. Andrews University, Scotland, and a Ph.D. degree in physics from the University of Pennsylvania. He has been a manager of R&D groups in electrophotographic materials, display materials and processes, organic electronics, and molecular electronics. Over the course of his research career, he has applied a wide variety of experimental techniques to the understanding of electronic, magnetic, and optical properties of primarily, but not exclusively, organic materials. He has experience in applying his knowledge of physics, materials science, chemistry, and processing to the technologies of electrophotographic printing, magnetic data storage, flat-panel displays, and memory devices. He has published more than 160 papers and holds 16 patents. He is a Fellow of the American Physical Society and a member of the Materials Research Society.

Chung H. Lam *IBM Research Division, Thomas J. Watson Research Center, P.O. Box 218, Yorktown Heights, New York 10598 (clam@us.ibm.com).* Dr. Lam received a B.S. degree in electrical engineering from the Polytechnic University of New York. He joined IBM in 1978 as a memory circuit designer. In 1984 he was awarded the IBM Resident Study Fellowship and received M.S. and Ph.D. degrees, both in electrical engineering, from Rensselaer Polytechnic Institute. He then assumed responsibilities in various disciplines of semiconductor research and development, including circuit and device design and process integration for memory and logic applications. He has been a member of the Technical Committee of the IEEE Non-Volatile Memory Workshop since 2001. He was named an IBM Distinguished Engineer in 2007. He has managed the Phase Change Memory Research Project since 2003.

Kailash Gopalakrishnan *IBM Almaden Research Center, 650 Harry Road, San Jose, California 95120 (kailash@us.ibm.com).* Dr. Gopalakrishnan received a B.S. degree in electrical engineering from the Indian Institute of Technology and M.S. and Ph.D. degrees in electrical engineering from Stanford University. Since July 2004, Dr. Gopalakrishnan has been a Research Staff Member at the IBM Almaden Research Center. His main research interests are novel logic and memory devices, device modeling and simulation, and nanoimprint lithography.

Rohit S. Shenoy *IBM Almaden Research Center, 650 Harry Road, San Jose, California 95120 (rsshenoy@us.ibm.com).* Dr. Shenoy received a B.Tech. degree in engineering physics from the Indian Institute of Technology and M.S. and Ph.D. degrees in electrical engineering from Stanford University. Since 2005, he has been employed at the IBM Almaden Research Center as a Research Staff Member. His current research interests include novel devices and process technology for ultrahigh-density memory and CMOS logic.