After Hard Drives—What Comes Next?

Mark H. Kryder and Chang Soo Kim

Department of Electrical and Computer Engineering, Data Storage Systems Center, Carnegie Mellon University,
Pittsburgh, PA 15213-3890 USA

There are numerous emerging nonvolatile memory technologies, which have been proposed as being capable of replacing hard disk drives (HDDs). In this paper, the prospects for these alternative technologies to displace HDDs in 2020 are analyzed. In order to compare technologies, projections were made of storage density and performance in year 2020 for both hard disks and the alternative technologies, assuming the alternative technologies could solve their remaining problems and assuming that hard drives would continue to advance areal density at a pace of about 40% per year, which would result in a two-disk 2.5-in disk drive that stores approximately 40 Terabytes and costs about $40. A major conclusion of the study is that to compete with hard drives on a cost per terabyte basis will be challenging for any solid state technology, because the ITRS lithography roadmap limits the density that most alternative technologies can achieve. Those technologies with the best opportunity have a small cell size and the capability of storing multiple bits per cell. Phase change random access memory (PCRAM) and spin transfer torque random access memory (STTRAM) appear to meet these criteria. PCRAMs are being marketed by at least one supplier and therefore appear to be closer to practical realization. On the other hand, STTRAMs would appear to have a performance edge assuming they, too, can be brought to market with multiple bits per cell. Although there are technologies that are not limited by the lithography roadmap and thus have greater areal density potential, they tend to be further from practical realization.

Index Terms—**Emerging alternative nonvolatile memory, hard disk drive, NAND flash.**

I. INTRODUCTION

MAGNETICALLY stored bits are theoretically stable in $L1_0$ FePt at densities approaching 100 Tb/in^2. With areal densities of today's drives around 500 Gb/in^2, hard disk drives (HDDs) are far from fundamental limits. The Information Storage Industry Consortium and its industrial sponsors from the HDD industry are targeting a demonstration of an areal density of 10 Tb/in^2 in 2015. Such a technology would enable over 7 TB to be stored on a single 2.5 inch disk, enabling a cost of the order of $3/TB for a two-disk 2.5 inch drive. Given the current 40% compound annual growth rate in areal density, this technology should be in volume production by 2020.

On the other hand, NAND flash memories have developed a significant presence in the nonvolatile memory (NVM) market and are now attempting to move into the computer storage market in the form of solid state drives (SSDs). Flash memories offer lower power consumption, faster read access time, and better mechanical reliability than HDDs; however, the cost per gigabyte (GB) for flash memories is nearly 10× that of magnetic storage. Moreover, flash memories face significant scaling challenges due to their dependence upon reductions in lithographic resolution as well as fundamental physical limitations beyond the 22 nm process node, such as severe floating gate interference, lower coupling ratio, short channel effects, and low electron charge in the floating gate. Thus, to replace HDDs, alternative NVM technologies that can overcome the shortcomings of NAND flash memories and compete on a cost per TB basis with HDDs must be found.

TABLE I
CHARACTERISTICS OF ALTERNATIVE NVM TECHNOLOGIES

Device Type	HDD	DRAM	NAND Flash	FRAM	MRAM	STTRAM	PCRAM	NRAM
Maturity	Product	Product	Product	Product	Product	Prototype	Product	Prototype
Present Density	400Gb/in^2 [7]	8Gb/chip [9]	64Gb/chip [10]	128Mb/chip	32Mb/chip	2Mb/chip	512Mb/chip	NA
Cell Size (SLC)	$(2/3)F^2$	$6F^2$	$4F^2$	$6F^2$	$20F^2$	$4F^2$	$5F^2$	$5F^2$
MLC Capability	No	No	4bits/cell	No	2bits/cell	4bits/cell	4bits/cell	No
Program Energy/bit	NA	2pJ	10nJ	2pJ	120pJ	0.02pJ	100pJ	10pJ [11]
Access Time (W/R)	9.5/8.5ms [8]	10/10ns	200/25ns	50/75ns	12/12ns	10/10ns	100/20ns	10/10ns [11]
Endurance/Retention	NA	10^{16}/64ms	10^5/10yr	10^{15}/10yr	10^{16}/10yr	10^{15}/10yr	10^9/10yr	10^{16}/10yr

Device Type	RRAM	CBRAM	SEM	Polymer	Molecular	Racetrack	Holographic	Probe
Maturity	Research	Prototype	Prototype	Research	Research	Research	Product	Prototype
Present Density	64Kb/chip	2Mb/chip	128Mb/chip	128b/chip	160Kb/chip	NA	515Gb/in^2	1Tb/in^2
Cell Size	$6F^2$	$6F^2$	$4F^2$	$6F^2$	$6F^2$	N/A	N/A	N/A
MLC Capability	2bits/cell	2bits/cell	No	2bits/cell	No	12bits/cell	N/A	N/A
Program Energy/bit	2pJ	2pJ	13pJ	NA	NA	2pJ	N/A	100pJ [12]
Access Time (W/R)	10/20ns	50/50ns	100/20ns	30/30ns	20/20ns	10/10ns	3.1/5.4ms	10/10us
Endurance/Retention	10^5/10yr	10^9/Months	10^9/days	10^4/Months	10^5/Months	10^{16}/10yr	10^5/50yr	10^5/NA

II. EMERGING NONVOLATILE MEMORY TECHNOLOGIES

In this paper, thirteen alternative NVM technologies are evaluated with respect to density, device performance, and likelihood of success in 2020. These technologies are listed in Table I along with HDDs, DRAM, and NAND Flash, which are included for comparison purposes. The cell sizes of all memory technologies in units of minimum feature size F were projected based upon the Emerging Research Devices (ERD) chapter of the 2007 International Technology Roadmap for Semiconductors (ITRS), which contains a tabulation of the recent experimental values as reported in technical [1]. Also indicated in Table I is whether the technology has the potential of storing multiple bits per cell (MLC capability) and an estimate of how many bits/cell might be achieved. The values of the other parameters such as program energy/bit, write and read access time and endurance/retention are based on an analysis of recently published technical papers and up-to-date product specifications as well as the ERD chapter of the 2007 ITRS [1]–[6].
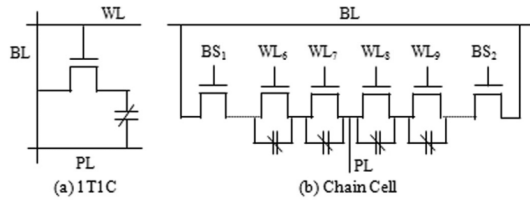
Fig. 1. FRAM cell architecture. (a) 1T1C, (b) Chain cell.

Fig. 2. Magnetic tunnel junction (MTJ) of MRAM.

Fig. 3. Cell structures of MRAM and STTRAM.

A. Ferroelectric RAM (FRAM)

In FRAM devices, common ferroelectric materials such as $Pb(Zr_x, Ti_{1-x})O_3$ (PZT), $SrBi_2,Ta_2O_9$ (SBT) and $(Bi, La)_4,Ti_2O_{12}$ are used to form a ferroelectric capacitor (FeCap), characterized by two remanent reversible polarization states. The cell structure is similar to a DRAM cell with the exception of the plateline (PL) that enables the switching of the FeCap, as shown in Fig. 1. The chain FRAM architecture, in which both contacts of each capacitor in the chain FRAM cell are shared with that of the adjacent capacitor, offers a smaller cell size of 6 F^2 than the conventional One Transistor and One Capacitor (1T1C) FRAM [13]. FRAM does not offer MLC capability. Ramtron International Corp. is offering FRAM products of 1 Mb, 2 Mb, and 4 Mb capacities. In 2009, Toshiba demonstrated a 0.252 μm^2, 128 Mb chain FRAM in a 4 M 130 nm process [14].

B. Magnetic RAM (MRAM)

MRAMs utilize the magnetization direction in the free layer of a two-layer magneto-resistive structure for data storage and the resultant resistance difference for information readout as shown in Fig. 2. There are three different physical effects for the realization of MRAMs: anisotropic magneto-resistance (AMR), giant magneto-resistance (GMR), and tunneling magneto-resistance (TMR). Both AMR cells and GMR cells consist of all-metal structures and result in low resistance change, which is not attractive for high density memories. On the other hand, TMR cells consist of a magnetic tunnel junction (MTJ) with two ferromagnetic layers separated by a thin dielectric and have recently been shown to exhibit TMR effects as high as 220% at room temperature [15]. The cell size of MRAMs is 20 F^2 and although MLC capability exists, no MLC MRAM product has been commercialized. In 2006, Freescale began shipping 4 Mb toggle MRAM chips with MgO-based TMR materials for use in cache buffers and configuration storage memories [16], and in 2009, NEC demonstrated a 32 Mb MRAM with 12 ns access time using two transistors and one MTJ per bit [17].

C. Spin Transfer Torque RAM (STTRAM)

STTRAMs have been investigated to solve the high cell writing current and large cell size problems posed by MRAMs. The MTJ structure of STTRAMs has two ferromagnetic layer and an MgO-based tunneling barrier layer in which thickness is controlled to less than 1 nm. Switching MTJ states from antiparallel or "1" to parallel or "0" and vice versa is performed by running a polarized electron current from the top to the bottom of the MTJ and vice versa [18]. The polarized current transfers angular momentum to the spins in the magnetic free
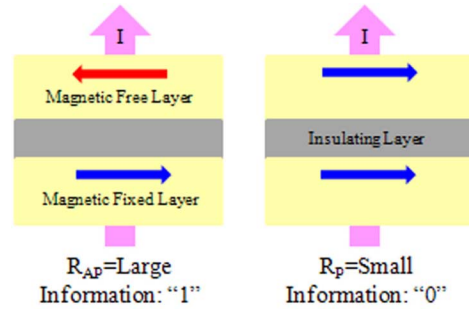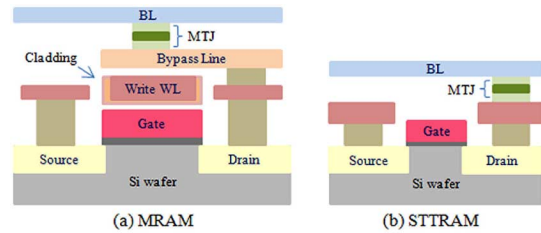
layer causing it to switch. The read operation of STTRAMs is basically the same as that of MRAMs. The use of polarized electron spins, rather than the field generated by the current, for writing, means that STTRAM cells do not require a write word line, bypass line or cladding in the cell structure as shown in Fig. 3, [19]. The cell size in SLC STTRAMs could be as small as 4 F^2 and 4 bit MLC STTRAMs may be possible with stacked MTJ structures [20]. In 2007, Hitachi demonstrated a 2 Mb STTRAM with a MgO tunneling barrier using 0.2 μm processing [21]. Grandis Inc. has recently begun prototyping 300-mm wafers in their MTJ fabrication facility [22].

D. Phase Change RAM (PCRAM)

PCRAMs utilize a reversible phase change between the amorphous and the crystalline states of a chalcogenide glass ($Ge_2Sb_2Te_5$, or GST) to produce a reversible resistance change in the cell as shown in Fig. 4, [23]. There are three types of switches used to select in PCRAMs. Diode switch PCRAMs offer cell sizes as small as 5 F^2 without loss of current driving capability, and prototypes have recently been made. Large resistance ON/OFF ratio of PCRAMs offers 4 bit MLC capability. In 2007, a fully functional 512 Mb SLC PCRAM chip with a 0.047 μm^2 cell (5.8 F^2) was demonstrated by Samsung using 90 nm technology [24]. In December 2008, Numonyx Inc., an Intel-ST Microelectronics joint venture, began commercial shipments of the industry's first 128 Mb MLC (2 bit) PCRAM products using a 90 nm process [25].

E. Carbon Nanotube RAM (NRAM)

A schematic diagram of a cell array of NRAM devices, in which carbon nanotube (CNT) arrays are suspended across a gap between the source and drain electrodes either with or without contact with the electrode depending on the voltage of the electrode, is shown in Fig. 5. Van der Waals forces make NRAMs nonvolatile by holding the CNTs in the bent position
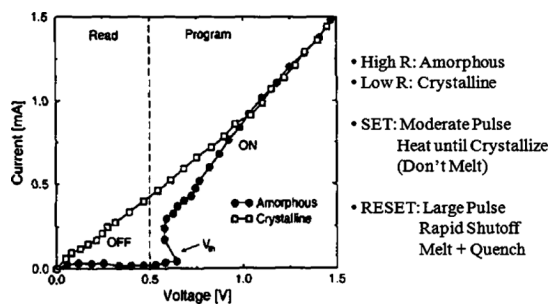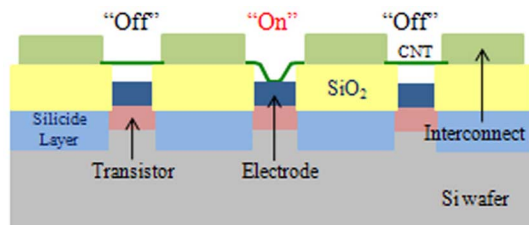
Fig. 4. I-V characteristics of PCRAM.

Fig. 5. Cell structure of NRAM.

Fig. 6. Device structure of MEMS based probe memory.

Fig. 7. Recording schematic of holographic memory.

that represents the "1" state until a pull-out voltage is applied to turn the device back to the "0" state [26]. The nonvolatility of NRAMs can be enhanced by increasing the length of suspended CNT arrays, decreasing the gap between CNT arrays and the gate or by selecting the surface layer to enhance the van der Waals interaction. A stronger van der Waals interaction causes an increase of pull-out voltage and a decrease of pull-in voltage. For optimum operation, it is critical to carefully control the pull-out and pull-in voltages in NRAMs. The cell size of NRAMs can be as small as 5 F^2, but there is no approach to creating multi-bit cells, as indicated in Table I. Nantero Inc. has recently fabricated and successfully tested a NRAM memory using 22 nm technology [11].

F. Probe Memory

Probe memories utilize cantilevers like those in atomic force microscopes (AFMs) to write and read information on the medium surface. The first probe memory technology was the IBM Millipede system that used a 2-D array of AFM cantilevers as thermo-mechanical scanning probes to create indentations where "1's" were to be written and no indentations where "0's" were to be written on a polymer-based medium [27]. Another approach is to use ferroelectric storage media that offer densities over 1 Tb/in^2 [28]. Yet another approach is to use an electro-thermal recording process, in which a phase change material like that in PCRAMs is used as a medium with a conductive bottom electrode and a suitable capping layer. The medium is altered by the flow of electrical current from the probe through the medium toward the bottom electrode [29]. In principle the medium of a probe memory can be featureless and the heads can be made by processes not strictly limited by lithography as indicated by the absence of a cell size in Table I. Nanochip Inc., a US start-up company, is working on a probe memory device using micro electro mechanical systems (MEMS) as shown in Fig. 6 [30], and says that it will
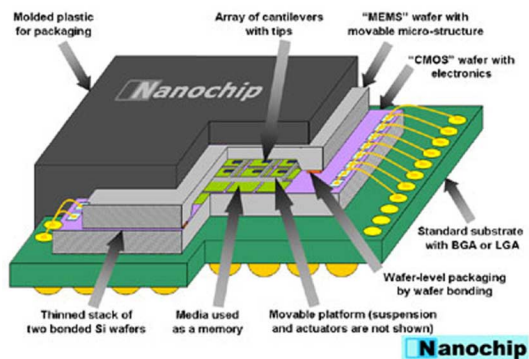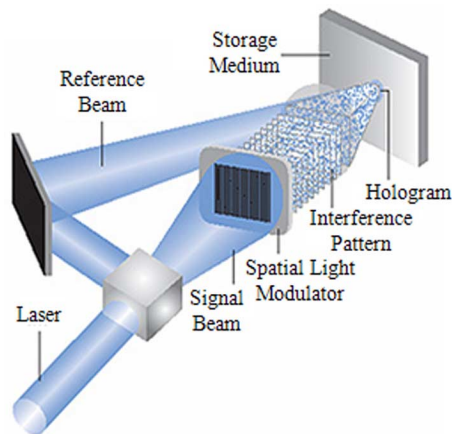
demonstrate exponentially higher density storage products with a cost/GB significantly below that of flash memories in 2010 [31].

G. Holographic Memory

Holographic memory stores and retrieves more than a million bits of data with a single flash of light. In order to do this, digital data is imposed onto an optical wavefront, stored holographically with high volumetric density, and then extracted from the retrieved optical wavefront as shown in Fig. 7 [32], [33]. Areal density is limited by the wavelength of the laser light used, rather than lithographic resolution. Holographic storage is currently being pursued as an archival storage technology with the ability to preserve data without degradation for more than 50 years. InPhase Technologies brought the first holographic storage product (called "Tapestry™") to market. It offers Write Once, Read Many times (WORM) functionality with a capacity of 300 GB and data transfer rates of 20 MB/s [34].

H. Copper Bridge RAM (CBRAM)

CBRAMs have been called programmable metallization cells (PMCs) and are comprised of a solid state electrolyte in which mobile metal ions move to generate a conductive bridge between two electrodes under the influence of an electric field as shown in Fig. 8 [35]. CBRAMs use a Ag doped Ge_xS_{1-x} base as a solid electrolyte and Ag top electrodes, where Ag is incorporated into the base using ultraviolet diffusion. The "ON" state
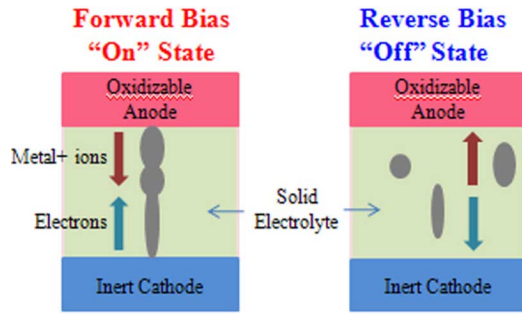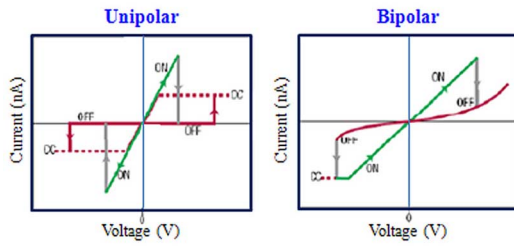
Fig. 8. Switching mechanism of CBRAM.

Fig. 10. 3-D structure of racetrack memory.

Fig. 9. Unipolar and bipolar switching of RRAM.

Fig. 11. Schematic of single electron transistor.

is achieved by applying a large forward bias voltage to the oxidizable anode causing redox reactions that drive silver ions in the base glass. This causes a conductive bridge to form between the electrodes [36]. By applying a reverse bias, the number of silver clusters is reduced and the conductive bridge is broken, creating the "off" state. CBRAMs can reproducibly switch between high and low resistive states (ON/OFF resistance $\approx 10^6$). CBRAM offers a cell size of 6 F^2 as well as MLC capability. In 2006, Altis Semiconductor, a French joint-venture between Infineon and IBM, demonstrated a 2 Mb 1T1CBJ CBRAM test chip using a 90 nm process [37].

I. Resistive RAM (RRAM)

RRAM memory cells are capacitor-like structures that exhibit a resistive switching phenomenon in transition metal oxides as shown in Fig. 9. There are two types of switching behaviors of RRAM devices: unipolar switching where the switching direction depends on the amplitude of the voltage, and bipolar switching where the switching direction depends on the polarity of the voltage. A variety of models for resistive switching have been suggested, including metallic filament formation, electron-trapping/defect-controlled switching, and crystalline-to-amorphous phase transition. Doped-$SrZrO_3$, ferroelectric $PbZrTiO_3$, and ferromagnetic $Pr_{1-x}Ca_xMnO_3$ (PCMO) as well as a variety of binary metal oxides such as Cu_xO, NiO, TiO_x, ZrO_x, and HfO_x have been investigated [38]. RRAM cell size is 6 F^2, and MLC operation is possible. Spansion demonstrated a 64 Kb memory test array with the Cu_xO MIM structure using 0.18 μm technology in 2005 [39].

J. Racetrack Memory

In a racetrack memory, information is stored in the form of domain walls along magnetic racetracks on a silicon wafer as shown in Fig. 10 [40]. T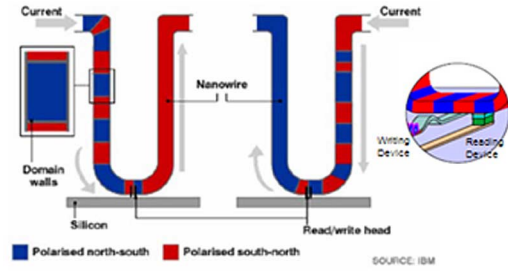he domain wall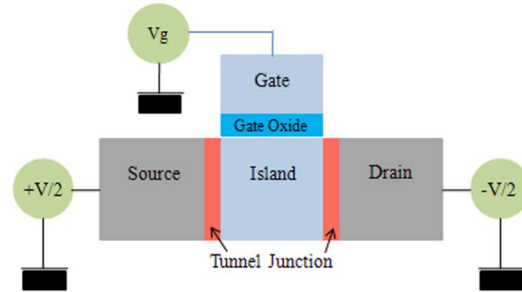s are caused to move in synchronism around the racetracks by applying spin-polarized current pulses. Information is written and read by a read/write head located at specific locations along the shift registers. This technology accommodates multiple magnetic domain walls per racetrack, and the spacing between consecutive magnetic domain walls that defines the bit length is controlled by pinning sites fabricated along the racetrack. By storing multiple bits in the vertically oriented racetracks, the storage can be 3-D enabling multiple bits per read/write cell. In Table I, 12 bits were assumed in each racetrack. In 2008, researchers at the IBM Almaden Research Center successfully demonstrated racetrack memories that were made of an array of magnetic permalloy nanowires that were 1–10 μm in length and 100 nm or less in diameter, proposing that it could potentially hold 100 times more data than flash memory does today [41].

K. Single Electron Memory (SEM)

SEMs utilize one-electron-precision charge transfer to a quantum dot as shown in Fig. 11 [42]. In a quantum dot, or 3-D island, an electron is confined electrostatically and controlled with integer electron precision. The injection of each electron into a quantum dot is performed across a tunneling barrier and controlled by a separate gate electrode via the Coulomb blockade effect. SEMs were initially able to operate only at temperatures below 4.2 K., but room temperature operation was recently demonstrated [43]. Although SEMs can store a single electron, their cell size is still 4 F^2 and they do not offer MLC capability. In 1998, Hitachi demonstrated a 128 Mb (8k $\times$ 8k $\times$ 2) memory chip using a 0.25 μm process and a cell size of 0.145 μm^2/bit [44].

L. Molecular Memory

A wide range of molecular memories have been investigated as possible building blocks of memory cells. One approach is to
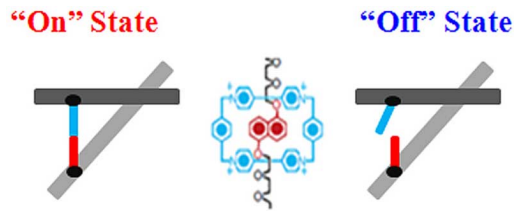
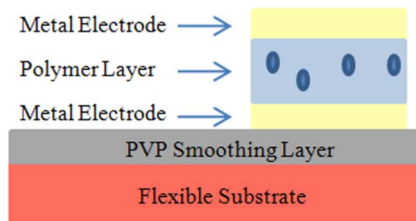Fig. 12.   Structure of two-terminal molecular memory.

Fig. 13.   Device structure for polymer memory.

utilize a two-terminal molecular switch tunnel junction, where a molecule switches from a stable isomer structure to another metastable isomer with a different conductivity at a specific voltage and remains in the latter state until another voltage pulse is applied to return it to the original isomer state as shown in Fig. 12 [45]. Those two states correspond to the "ON" and "OFF" states of the molecular memory device. The cell size is a moderate 6 F^2 and there have been no reports of MLC demonstration. In 2007, researchers at Caltech successfully demonstrated a 160 Kb molecular memory device with a density of 10^{11} bits cm^{-2} [46].

M. Polymer Memory

A polymer memory device is a two-terminal bistable device in which a polymer film is sandwiched between two metal electrodes, as shown in Fig. 13. In order to write a bit, a voltage is applied across the structure, causing charge transfer between the metal atoms and the polymer compounds, leading to a change in electrical conductivity [47]. The erase operation is performed by reversing the voltage. This switching phenomenon could in principle be expanded to three-terminal bistable devices with two organic layers and a middle discontinuous metal layer, sandwiched between two metal electrodes in order to achieve 2 bit polymer memory devices. The SLC cell size is 6 F^2 as shown in Table I. A 16-byte polymer memory array on a plastic substrate was successfully demonstrated in 2007 and Advanced Micro Devices is reportedly working on developing new products for flexible electronic applications [48].

III. EVALUATION CRITERIA

Since HDDs are expected to have an areal density of 10 Tb/in^2 in 2020 and to be within an order of magnitude of their ultimate limit, and since flash memories are expected to reach their areal density limit before 2020, it was decided to compare the emerging technologies on the basis of their potential for replacing flash memories and competing with HDDs in the 2020 timeframe. Although there are many different possible

bases for comparison, five criteria were selected as being most significant. They were density (which is determined by the cell size divided by the number of bits per cell), power efficiency, write and read access time, endurance, and retention time. All of these are listed in Table I and the values shown are projections of what it is believed the technology could achieve in the 2020 timeframe, assuming technological hurdles are overcome and the technology reaches volume production.

Density is viewed as the most important factor in determining whether a new NVM technology will be successful or not, because it relates directly to cost/GB and in the HDD marketplace, cost/GB has always been substantially more important than other performance parameters. To compare cost/GB, the approach used here was to assume that, to first order, cost/GB would scale in proportion to (density)$^{-1}$, which is cell size divided by the number of bits per cell. This assumption is generally valid so long as the cost of wafer processing and packaging does not vary drastically.

The next most important performance parameter is believed to be power efficiency. Power is critical in mobile devices, because battery life is limited by the power the device draws. Although power has traditionally not been a large concern in data centers, today, power and cooling for data centers is a significant fraction of their expense. For equivalent capacity, NAND flash based SSDs draw about half the power of HDDs. However, for the emerging NVM technologies, SSDs are not available. Consequently the various NVM technologies were compared on the basis of their program energy per bit.

Access time, or the time interval between the write/read request and the writing or availability of the data, appears to be the third most important criterion. These parameters are typically more important in data center applications than in consumer applications.

Endurance, or the number of times a bit may be rewritten, and retention, or the length of time a bit remains stable, are generally less important, so long as they meet minimum criteria. Although systems today require endurance of the order of 10^{15} in the file access table (FAT) in a HDD, single level cell (SLC) NAND Flash memories with endurance of the order of 10^5 cycles are made to work in systems by moving the FAT file to new locations before the 10^5 cycle limit is reached. This requires some added complexity in the controller and a small sacrifice in performance, but these have not proven to be major issues. Retention is generally required to be at least five years for commercial products, but for archival applications 50 or even 100 years may be required.

IV. ASSESSMENT

In order to obtain an overall comparison of the various technologies with one another, the data in Table I are plotted on spider diagrams in Fig. 14. The data are plotted in the spider diagrams on a logarithmic scale, where the value 1 indicates the worst performance among all these memory technologies and the value 5 indicates the best performance of all these memory technologies.

Holographic memory, probe memory and to some extent, racetrack memory, are not limited in the same way as the other
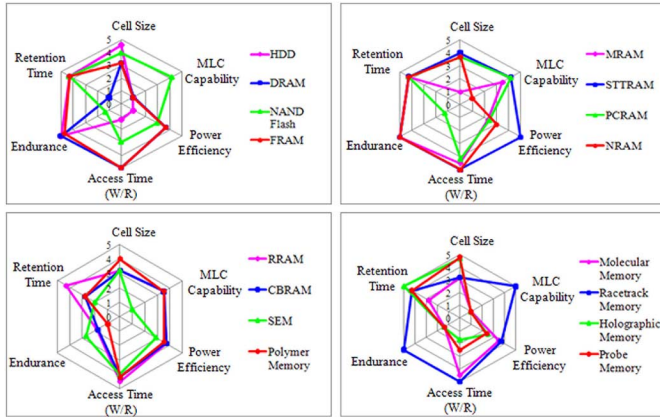
Fig. 14.   Performance assessment of memory technologies.

technologies by lithographic progress. Hence, they offer higher density potential than most of the other technologies.

Holographic memory offers high density and excellent retention, but relatively poor endurance, access time and power efficiency as shown in Fig. 14. Today it only offers WORM functionality and, although advocates of the technology claim it could be made rewritable, there have been no reports of practical rewritable devices. The main market that holographic memory has attempted to address is the archival market, which requires a 50 year storage lifetime on a low cost medium. In addition, holographic memory requires a relatively complex and expensive opto-electronic system for recording and readback, resulting in the cost of the recorders being 10–100× more than HDDs. Media stability and a high bit error rate that requires extensive correction are also issues.

Probe memories can in principle record on a featureless medium using probe heads that are not limited by lithography. Thus, they have very high density potential and have also exhibited good retention; however, this depends upon the medium selected. On the other hand, they have thus far had access times not much better than HDDs, require relatively high power and endurance which may be adequate, but could also be an issue. Most companies that have pursued this technology have used MEMS, which thus far has not proven to offer low power or low cost and which has access times comparable to hard drives. In order to achieve competitive data rates, many heads must be accessed in parallel, which drives up total cost and power. Probe memories thus offer potential to make an ultra-high density storage device with moderate performance; however, much remains to be done to make it a practical technology.

Racetrack memories offer the possibility of making 3D memories and thus potentially offer many bits per cell, even while using lithography to define the structures. They also potentially offer low power, fast access time, good endurance and good retention as shown in Fig. 14. However, making a 3-D working racetrack memory remains to be demonstrated, and it is not clear how the bit cells in the racetracks will be defined without using lithography and whether it will, in fact, be possible to move all the data in a racetrack in synchronism from cell to cell without error. There have been previous attempts to make devices that

use shift registers for wall motion, such as the Cross-tie memory [49] and the Bloch line memory [50], but they have failed to be brought to market. In addition, the high current density required to drive the domain wall motion, the high aspect ratio structures that are required in the racetracks and developing a fabrication process, which is compatible with CMOS processing, are serious challenges. Racetrack memory is thus an interesting candidate, offering what might be judged to be tremendous potential, but requiring what appears to be a large effort to make work.

The advantages of MRAM devices are their relatively fast access times, excellent endurance and long retention times as seen in Fig. 14. Additionally, these are achieved with straightforward CMOS back-end-of-line (BEOL) comparability. As a result of these advantages, MRAMs have met with some success in embedded or system-on-chip (SoC) applications. However, the relatively large cell size due to complex CMOS BEOL integration and their poor power efficiency due to high cell write current are fundamental obstacles to scaling the device to sufficiently high density and low cost to compete with flash memories or HDDs. One approach to decrease the large cell size down to approximately $4 F^2$ is to utilize a cross-bar architecture with no isolation transistor, but this limits the speed of the technology, and does not get around the fact that the current required to switch the MRAM cell does not scale. On account of this, it is highly unlikely that MRAMs will be taken to much higher densities.

FRAMs offer relatively low power consumption, fast access time, excellent endurance and excellent retention as shown in Fig. 14. This technology has easy complementary CMOS BEOL comparability, and has been used for computer systems and embedded applications. On the other hand, FRAMs have relatively low potential for scaling down their cell size, since it is difficult to reduce the 3-D FeCap without sacrificing device performance and reliability due to both degradation of the cell signal margin and difficulty in reducing the parasitic capacitance. Hence, it is doubtful that FRAM will ever be a serious competitor of flash memories or HDDs.

The attractiveness of SEM devices is the intriguing idea that one could store information in something as small as a single electron, but the practical implementation of the device structure offers a cell size no smaller than $4 F^2$ as indicated in Table I. They also offer potential for fast access time and reasonable endurance, but require relatively high power and the worst retention time of all the NVM technologies discussed here (particularly, if room temperature operation is required) as shown in Fig. 14. In addition, cell-to-cell variation is a serious issue for the manufacture of SEM devices. Although the time may come when the ability to write and read a single electron is critical to achieving higher densities, it seems unlikely to occur in the 2020 timeframe addressed here.

Molecular memory devices offer relatively good power efficiency and access time, but only moderately small cell size ($6 F^2$), no MLC capability and relatively poor endurance and retention times as shown in Fig. 14. Other challenges are the compatibility of molecular devices with CMOS processing and also the functional assembly and interconnection of nanoscale molecules in a device having capacities in the gigabyte to terabyte range. Thus, molecular memory devices are also unlikely to be products in the 2020 timeframe.

Polymer memory devices have higher areal density potential than molecular memory devices because of their MLC potential and offer relatively good power efficiency and access time, but only moderately small cell size and relatively poor endurance and retention times as shown in Fig. 14. A major attraction for polymer memory devices is the fact that they may be able to be fabricated on flexible substrates, making them potentially attractive for flexible electronic applications such as RFID tags, e-signage, and e-paper. However, given their poor endurance and retention times, it seems unlikely that polymer memories will compete with flash memories or HDDs.

CBRAMs exhibit high potential with regard to cell size, power efficiency and access time, and reasonable endurance. They can also achieve MLC technology by controlling the programming current [51], but have had difficulty with retention time as shown in Fig. 14. Though they are proposed for a variety of embedded applications, it seems unlikely that they will compete with the cost/GB or retention time of flash memories or HDDs in 2020 timeframe.

RRAMs offer high density potential due to small cell size and MLC capability. They also offer good power efficiency, fast access times and reasonable endurance as shown in Fig. 14. In addition, they have good thermal budget tolerances and easy CMOS BEOL comparability, making it possible to embed this technology in other devices. On the other hand, RRAM devices have not demonstrated robust endurance. Only a NiO memory cell was demonstrated to have endurance of over 10^6 [52]. In addition, there have been large cell-to-cell and die-to-die variations in transistor characteristics, as well as high leakage currents and high reset currents. Unless these problems can be solved, this technology is unlikely to be brought to market in the 2020 timeframe.

NRAMs have excellent potential for achieving fast access times, high endurance, long retention, and good power efficiency, but offer moderate cell size (5 F^2) with no MLC capability as shown in Fig. 14. They thus could be used for both standalone and embedded memory applications such as CPU cache or for replacing DRAM and NOR flash memories, but will probably not compete with NAND flash memories. Moreover, using CNTs in CMOS fabs causes metallic contamination, and there are significant difficulties in controlling the diameter of CNTs and in uniformly patterning suspended CNT ribbons. Thus, although NRAMs have a potential to become products, it seems unlikely that they will compete with flash memories or HDDs.

PCRAMs offer small cell sizes with multi-bit per cell technology, relatively fast access time, good retention time and reasonable endurance comparable to that of NAND flash, but require somewhat higher power than most other technologies, as shown in Fig. 14. They are a relatively mature NVM technology and are anticipated to replace high density NOR flash memories. However, to compete with NAND flash memories, their RESET current, and, correspondingly their power requirements must be reduced.

Finally, STTRAMs appear to potentially offer superior power efficiency, access time, endurance, and retention time as shown in Fig. 14. Thus, they currently appear to have potential to replace NOR flash memories and even DRAMs. If they can be made to store multiple bits per cell, then they could also offer extremely small cell size and low cost, in which case they could also become a viable candidate for replacement of NAND flash and perhaps even HDDs.

V. CONCLUSION

Assuming HDDs continue to progress at the pace they have in the recent past, in 2020 a two-disk, 2.5-in disk drive will be capable of storing over 14 TB and will cost about $40. Long before 2020, flash memory technology, on the other hand, will reach limits that will prevent its continued scaling. Consequently there is today a lot of interest in alternative NVM technologies that could replace flash memories and perhaps ultimately displace HDDs. In this paper, thirteen of them were compared in terms of their potential density and performance in 2020.

Taking achievable density as the most important parameter and factoring in the other performance characteristics, racetrack memories appear to offer considerable potential; however, racetrack memories rely upon synchronous motion of domain walls in shift registers, something which has proven to be unreliable in a number of other devices, is far from practical application and does not appear to have the critical mass of researchers working on it to solve all the critical problems. Probe based memories offer the potential of very high density, because they can use a featureless medium with heads that may be able to be made with processes that are not severely constrained by lithographic tools. On the other hand, their performance lags that of other technologies and, like racetrack memories are far from practical application with a small number of researchers pursuing them. Holographic memories have high-density/low-cost potential, but have only been demonstrated as a write-once technology, which is an entirely different market. MRAMs and FRAMs are the most expensive technologies discussed here with costs similar to that of DRAMs and are therefore not candidates to replace either flash memories or HDDs. Polymer memories, molecular memories and SEMs all have lower densities than required to compete with flash memories and retention times that are too short to be practical for storage technologies. NRAMs appear to meet the requirements of future NVM technologies but their projected density and therefore cost is not sufficiently competitive to replace even flash memories. RRAMs, CBRAMs, STTRAMs, and PCRAMs all have small cell size and potential MLC capability, which could enable densities and cost/TB comparable to that of disk drives. Of these, PCRAMs are the most mature, already being offered as products, while STTRAMs appear to offer superior performance.

REFERENCES

[1] [Online]. Available: http://www.itrs.net/Links/2007ITRS/2007_Chapters/2007_ERD.pdf

[2] S. A. Wolf *et al.*, "Spintronics-A retrospective and perspective," *J. Res. Develop.*, vol. 50, no. 1, Jan. 2006.

[3] [Online]. Available: http://www.sciencemag.org/cgi/content/full/320/5873/190/DC1/

[4] G. W. Burr *et al.*, "Overview of candidate device technologies for storage-class memory," *J. Res. Develop.*, vol. 52, no. 4/5, July/Sept. 2008.

[5] J. E. Brewer *et al.*, "Memory technology for the post CMOS era," *IEEE Circuits Devices Mag.*, Mar./Apr. 2005.

[6] K. Galatsis *et al.*, "Emerging memory devices," *IEEE Circuits Devices Mag.*, vol. 22, pp. 12–21, May/June 2006.

[7] [Online]. Available: http://www.pcworld.com/reviews/product/44137/review/caviar_green_2tb_wd20eads.html/

[8] [Online]. Available: http://www.seagate.com/www/en-us/products/desktops/barracuda_harddrives/

[9] U. Kang *et al.*, "8 Gb 3D DDR3 DRAM using through-silicon-via technology," *ISSCC*, vol. 52, pp. 130–131, Feb. 2009.

[10] C. Trinh *et al.*, "A 5.6 MB/s 64 Gb 4 b/Cell NAND flash memory in 43 nm CMOS," *ISSCC*, vol. 52, pp. 246–247, Feb. 2009.

[11] R. F. Smith *et al.*, "Carbon nanotube based memory development and testing," in *Proc. IEEE Aerospace Conf.*, Mar. 2007.

[12] [Online]. Available: http://projects.exeter.ac.uk/dsnet/pdf/WhiteBook_Ver2006.pdf/

[13] H. Kanaya *et al.*, "A 0.602 μm^2 nested chain cell structure formed by one mask etching process for 64 Mbit FeRAM," in *Symp. VLSl Technology Dig. Tech. Papers*, 2004.

[14] H. Shiga *et al.*, "A 1.6 GB/s DDR2 128 Mb chain FeRAM with scalable octal bitline and sensing schemes," in *ISSCC, Session 27*, Feb. 2009, pp. 464–465.

[15] S. S. P. Parkin and C. Kaiser, "Giant tunnelling magnetro-resistance at room temperature with MgO 100 tunnel barriers," *Nature Mater.*, 2004.

[16] B. N. Engel *et al.*, "A 4-Mbit toggle MRAM based on a novel bit and switching method," *IEEE Trans. Magn.*, vol. 41, pp. 132–136, 2005.

[17] R. Nebashi *et al.*, "A 90 nm 12 ns 32 Mb 2T1MTJ MRAM," in *ISSCC, Session 27*, Feb. 2009, pp. 462–463.

[18] M. Hosomi *et al.*, "A novel nonvolatile memory with spin torque transfer magnetization switching: Spin-RAM," in *IEDM Tech. Dig.*, Dec. 2006, pp. 473–476.

[19] [Online]. Available: http://www.embedded.com/columns/technicalinsights/197006966?_requestid=1242363/

[20] I.-G. Baek, J.-E. Lee, S.-C. Oh, K.-T. Nam, and J.-H. Jeong, "Memory Devices Including Multibit Memory Cells Having Magnetic and Resistive Memory Elements and Related Methods," U.S. Patent 2008/0180989 A1, July 31, 2008.

[21] T. Kawahara *et al.*, "2 Mb spin-transfer torque RAM (SPRAM) with bit-by-bit bidirectional current write and parallelizing-direction current read," in *IEEE Int. Solid-State Circuits Conf.*, 2007.

[22] [Online]. Available: http://www.grandisinc.com/pdf/Feb2_2009_Grandis_300_mm.pdf

[23] A. Pirovano *et al.*, "Electronic switching effect in phase-change memory cells," in *IEEE Int. Electron Devices Meeting, Tech. Dig.*, 2005, pp. 923–926.

[24] J. H. Oh *et al.*, "Full integration of highly manufacturable 512 Mb PRAM based on 90 nm technology," *IEDM Tech. Dig.*, pp. 49–51, 2006.

[25] [Online]. Available: http://numonyx.com/enUS/About/PressRoom/Features/Pages/PCMships.aspx/

[26] J. W. Ward *et al.*, "A non-volatile nano-electro-mechanical memory element utilizing a fabric of carbon nanotubes," in *Proc. 2004 Non-Volatile Memory Technology Symp.*, Nov. 2004, pp. 34–38.

[27] A. Knoll *et al.*, "Integrating nanotechnology into a working storage device," *Microelectron. Eng.*, vol. 83, pp. 1692–1697, 2006.

[28] K. Tanaka and Y. Kurihashi, "Scanning nonlinear dielectric microscopy nano-science and technology for next generation high density ferroelectric data storage," *Jpn. J. Appl. Phys.*, vol. 47, no. 5, pp. 3311–3325, 2008.

[29] C. D. Wright *et al.*, "Scanning probe-based phase-change terabyte memories," in *European\Phase Change and Ovonics Symposium*, 2008.

[30] [Online]. Available: http://www.nanochipinc.com/tech.htm/

[31] [Online]. Available: http://www.nanochipinc.com/NanoChip_Serie_C2_Announcement_pdf/

[32] [Online]. Available: http://www.inphase-technologies.com/technology/default.aspsubn=2_1/

[33] G. W. Burr, "Holographic storage," *Encyclopedia of Optical Engineering*, Sep. 2003.

[34] [Online]. Available: http://www.inphasetechnologies.com/downloads/pdf/products/2007Tapestry/ProductBrochure.pdf/

[35] C. Liaw *et al.*, "The conductive bridging random access memory (CBRAM): A non-volatile multi-level memory technology," in *Solid State Device Research Conf.*, Sept. 2007, pp. 226–299.

[36] M. N. Kozicki *et al.*, "Nanoscale memory elements based on solid-state electrolytes," *IEEE Trans. Nanotechnol.*, vol. 4, no. 3, May 2005.

[37] H. Hönigschmid *et al.*, "A non-volatile 2 Mbit CBRAM memory core featuring advanced read and program control," in *Symp. VLSI Circuits Dig. Tech. Papers*, 2006.

[38] A. Sawa, "Resistive switching in transition metal oxides," *Mater. Today*, vol. 11, no. 6, pp. 28–36, Jun. 2008.

[39] A. Chen *et al.*, "Non-volatile resistive switching for advanced memory applications," in *Proc. IEEE Int. Electron Devices Meeting*, 2005, pp. 746–749.

[40] M. Hayashi *et al.*, "Current-controlled magnetic domain-wall nanowire shift register," *Science*, vol. 320, Apr. 2008.

[41] S. S. P. Parkin *et al.*, "Magnetic domain-wall racetrack memory," *Science*, vol. 320, Apr. 2008.

[42] [Online]. Available: http://www.sematech.org/meetings/archives/other/20021028/18_Wasshuber_Electron%20Mem.pdf/

[43] L. Guo *et al.*, "A silicon single-electron transistor memory operating at room temperature," *Science*, vol. 275, 1997.

[44] K. Yano and T. Ishi, "128 Mb early prototype for gigascale single-electron memories," in *IEEE Int. Solid-State Circuits Conf.*, 1998.

[45] A. H. Flood *et al.*, "Whence molecular electronics?," *Science*, vol. 306, Dec. 2004.

[46] J. E. Green1 *et al.*, "A 160-kilobit molecular electronic memory patterned at 1011 bits per square centimeter," *Nature*, vol. 445, Jan. 2007.

[47] L. P. Ma *et al.*, "Organic electrical bistable devices and rewritable memory cells," *Appl. Phys. Lett.*, vol. 80, no. 16, p. 2997, Apr. 2002.

[48] H.-T. Lin *et al.*, "A 16-byte nonvolatile bistable polymer memory array on plastic substrates," in *IEEE Int. Electron Devices Meeting*, 2007.

[49] L. J. Schwee, "Stability conditions for neel walls and crosstie walls in thin magnetic films," in *AIP Conf. Proc.*, 1972, no. 10, p. 996.

[50] F. B. Humphrey and J. C. Wu, "Vertical Bloch line memory," *IEEE Trans. Magn.*, vol. MAG-21, no. 5, Sep. 1985.

[51] H. Hönigschmid *et al.*, "A non-volatile 2 Mbit CBRAM memory core featuring advanced read and program control," in *Symp. VLSI Tech.*, 2006, pp. 138–139.

[52] I. G. Baek *et al.*, "Highly scalable nonvolatile resistive memory using simple binary oxide driven by asymmetric unipolar voltage pulses," in *IEEE Int. Electron Devices Meeting*, Dec. 2004, pp. 587–590.