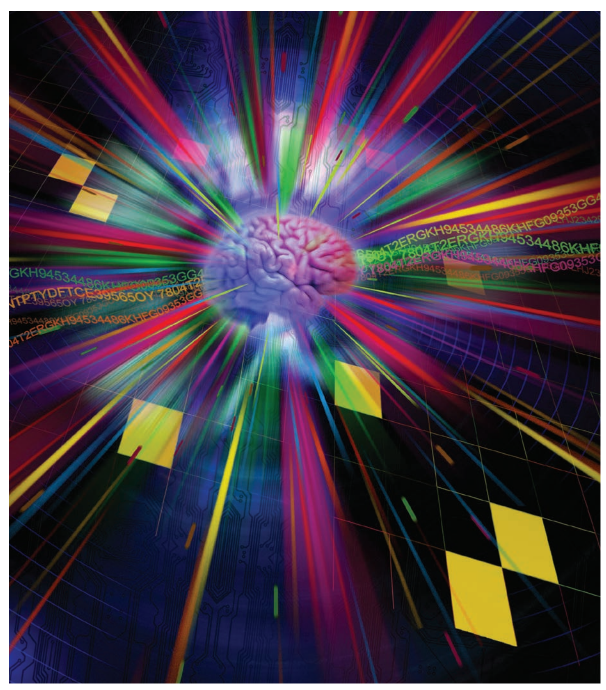


What Lies Ahead for Resistance-Based Memory Technologies?

Yoon-Jong Song, Gitae Jeong, In-Gyu Baek, and Jungdal Choi, ***Samsung Electronics***

Phase-change RAM, magnetic RAM, and resistive RAM offer strong scalability, speed, and power consumption advantages over conventional capacitance-based memory. Recent work shows the feasibility of mass producing these new devices and their suitability for next-generation technology.

Memory devices are rapidly approaching the technical limit of scalability, which is fueling interest in new technology. Conventional memory such as dynamic RAM (DRAM) and NAND flash exploit the capacitance from electrical charge stored in the memory cell. As that storage area shrinks, these devices must wrestle with the uncontrollable charge decrease in the reduced cell area.

Some researchers are trying to solve that problem by adapting new architectures such as vertical NAND in the current devices. Others are developing new resistance-based memory devices such as phase-change RAM (PRAM), magnetic RAM (MRAM), or resistive RAM (ReRAM). Relative to capacitance-based memory, these nonvolatile memory technologies are faster and more easily accommodate scaling to sub-20-nm nodes. Each has its own market segment: MRAM has the fastest speed, which places it as a nonvolatile alternative to DRAM. PRAM and ReRAM are faster than flash memory, which places them as contenders for the flash market.

To better understand possible memory choices, semiconductor manufacturers should familiarize themselves with the key characteristics and current process technology of these three new memory types.

PHASE-CHANGE RAM

Researchers have investigated PRAM as a promising device for standard interface or storage memory applications because of its nonvolatility, relatively fast operational speed, and superb scalability.[1,2] It basically applies two phases of chalcogenide materials: the amorphous phase with high resistance and the crystal phase with low resistance. In the amorphous phase, manufacturers melt the chalcogenide films through joule heating and quench the molten state into a frozen amorphous state. In the crystal phase, they supply enough heat and time to rearrange the chalcogenide element atoms.

Controlling the write voltage and time can reversibly transform the written data from the amorphous phase to the crystalline phase. The read data is the result of sensing the resistance difference between the two phases. Figure 1 shows PRAM's basic structure. Typically, $Ge_2Sb_2Te_5$ (GST) films are the phase-change material.

 Published by the IEEE Computer Society 0018-9162/13/$31.00 © 2013 IEEE

A critical electrical parameter is the reset current (I_{reset})—the current required to melt the chalcogenide film. The device must supply a larger current than the I_{reset}, which often translates to a large cell transistor size and thus a large chip. For PRAM to compete successfully with other memory devices, the I_{reset} must decrease as the device scales down. In one study of this reduction,[3] 20-nm PRAM showed an I_{reset} of about 80 microamperes (μA).

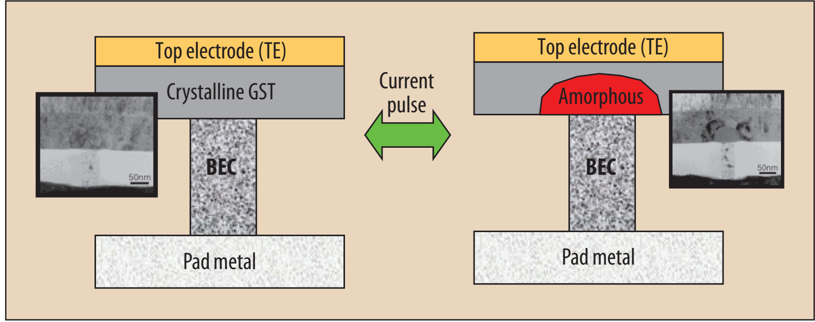


Figure 1. Phase-change RAM (PRAM) basic operation scheme. In the amorphous phase, characterized by high resistance, manufacturers melt the chalcogenide films through joule heating and quench the molten status into frozen amorphous status. The crystal phase, characterized by low resistance, rearranges the chalcogenide element atoms. BEC: bottom electrode contact; GST: $Ge_2Sb_2Te_5$.

Process technology

The integration process to control the I_{reset} is a central part of PRAM development and the source of schemes to accommodate down-scaled designs. The writing process involves joule heating the chalcogenide film, so reducing the contact area between the bottom electrode and the film decreases the I_{reset}. Fabricating a pillar-type bottom electrode contact (BEC) early in development is relatively simple, but as the design shrinks, a seam or void forms in the pillar structure because of the smaller contact area.

Advanced electrode schemes require the deposition of a very thin electrode inside the contact, which forms as a ring type. The contact area depends on the electrode's thickness, not the contact size, so a significant reduction is possible.[4] Controlling the electrode's resistivity further decreases the I_{reset}. Figure 2a shows transition behavior as a function of electrode resistivity. Enhanced joule heating increases electrode resistivity, which in turn decreases the I_{reset}. However, increased resistivity can also increase set resistance, which produces a narrow sensing margin. Thus, manufacturers must seek the appropriate tradeoff of I_{reset} and set resistance.

Controlling the chalcogenide film and its structure is also critical in reducing the I_{reset}. One scheme uses a confined cell structure in which GST films are deposited inside the electrode contact hole rather than on the electrode. As Figure 2b shows, the confined cell structure greatly reduces programming volume, which yields a low reset current. Doping the chalcogenide films also improves the electrical property. In one application,[5] doping increased the GST film resistance, thus improving heating efficiency and reducing the I_{reset}. Researchers attributed this high resistance to the interaction of the doping and GST elements. They found that, in the optimum doping concentration, GST films show a low I_{reset} without the cost of high set resistance.

Development status and future trends

Using these advanced cell technologies, we were able to develop a full integration scheme for the 20-nm technology node. Figure 3 shows the cross-sectional view of the fully integrated PRAM device using triple metallization and

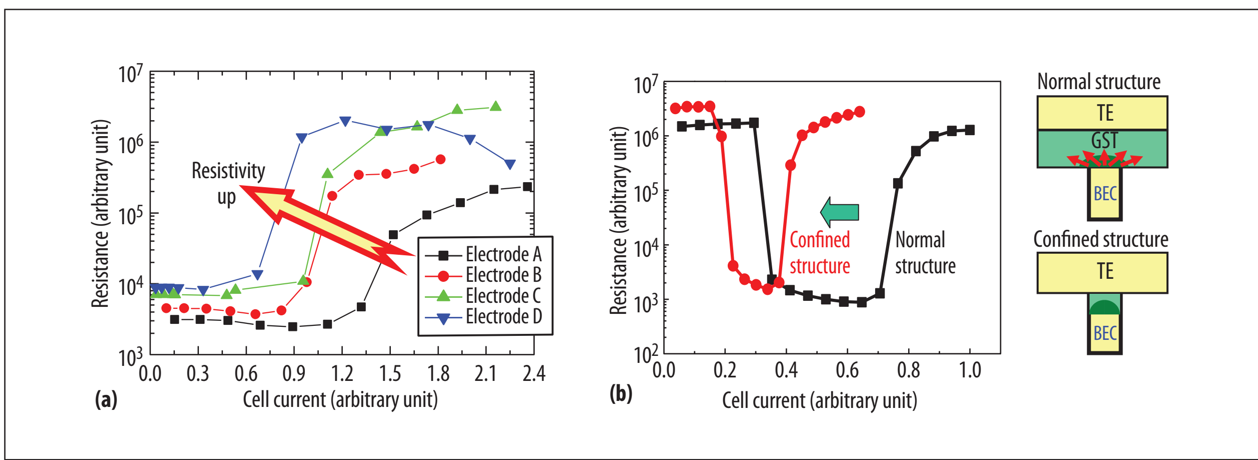


Figure 2. Decreasing the I_{reset} (a) transition curves as a function of electrode resistivity and (b) transition curves of normal and confined cell structures.

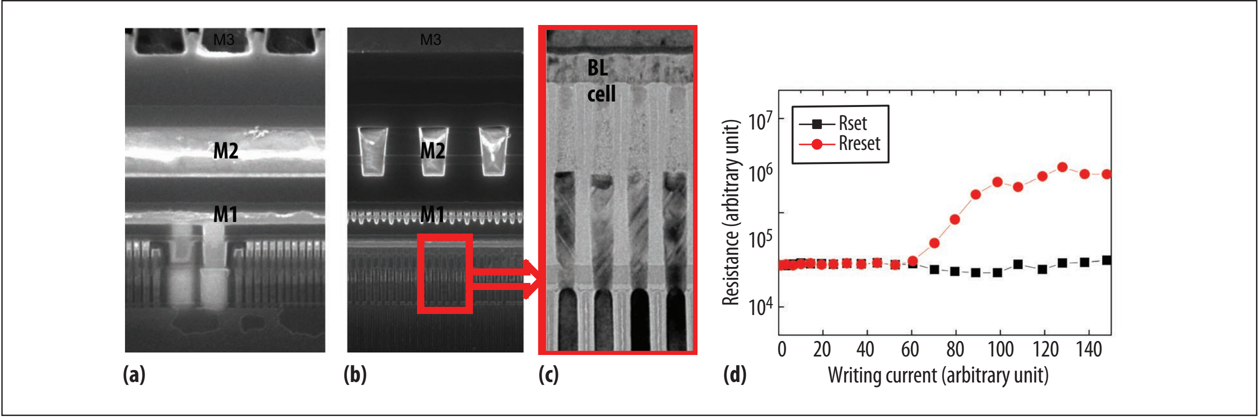


Figure 3. **Fully integrated phase-change RAM device (a) cross-sectional view of bit-line direction, (b) cross-sectional view of word-line direction, (c) cross-sectional view of transmission electron microscopy (TEM) cell image, and (d) transition curve of a 20-nm fully integrated GST cell.**

confined structure. Full integration produced a stable transition curve with the I_{reset} below 100 µA. Retention tests for the 20-nm cell array showed the feasibility of maintaining stored data for 10 years at 85°C.

Combining an innovative cell scheme and certain doping elements can further decrease the I_{reset}, but thermal disturbance will be a critical factor in developing a PRAM cell. The joule heating in narrow adjacent cells will require a cell structure with a special thermal-blocking scheme or doped phase-change materials that are extremely thermally stable.

MAGNETIC RAM

MRAM uses tunneling resistance that depends on the relative magnetization directions of ferromagnetic electrodes.[6-8] It is an attractive memory option because of its superior scalability, speed, and power consumption. Depending on the writing technique, MRAM is classified as either field switching or spin-transfer torque (STT-MRAM)—the latter type being of more interest because of its simple structure and excellent scalability.

As Figure 4 shows, the MRAM's memory cell stack comprises a pinned ferromagnetic electrode layer, tunneling oxide, and a free ferromagnetic electrode layer. When the free layer's and pinned layer's magnetizations are parallel, the magnetic tunnel junction (MTJ) stack exhibits low resistance (R_p). When the two magnetizations are not parallel, resistance is high (R_{ap}). The tunnel magnetoresistance (TMR) ratio, $(R_{ap} - R_p)/R_p$, defines the sensing margin between data 1 and data 0.

Magnetization can also be either in-plane (IP) or perpendicular to the plane (PP). Each magnetization direction has different requirements that need different magnetic materials. Recent attention is shifting to PP STT-MRAM because its lower switching current density and improved thermal instability produce better scalability.[9,10]

Two factors decide how far MRAM can shrink: the critical current density (J_c)—the current density for magnetic switching—and the thermal stability factor (Δ), which is closely correlated with retention. J_c should be as low as possible to fully exploit MRAM's low power consumption and high density advantages. In contrast, Δ should be as high as possible to guarantee reliable device functionality. Adjusting process parameters requires a tradeoff between the two.

Integration process

To illustrate the integration process, consider the scaling process and the magnetic switching properties of a 28-nm MTJ cell. A central part of integration is MTJ etching. Because of the MTJ cell's low angle-etching slope and its sidewall by-products, much work has concentrated on improving technology to support MTJ's patterning process. Figure 5a shows the cross-sectional TEM image of a 28-nm CoFeB-base MTJ cell using an advanced etching technique.[11] This process produced relatively steep MTJ cells without any undesirable side effects. Figure 5b shows the TMR value and interfacial perpendicular

Read current
(a)
(b)
Ferromagnetic electrode (free layer)
Tunneling oxide
Ferromagnetic electrode (pinned layer)

Figure 4. **Magnetic RAM (MRAM) basic operational scheme. (a) Parallel magnetization between the free and pinned layers results in low resistance (R_p). (b) Magnetization that is not parallel yields high resistance (R_{ap}).**

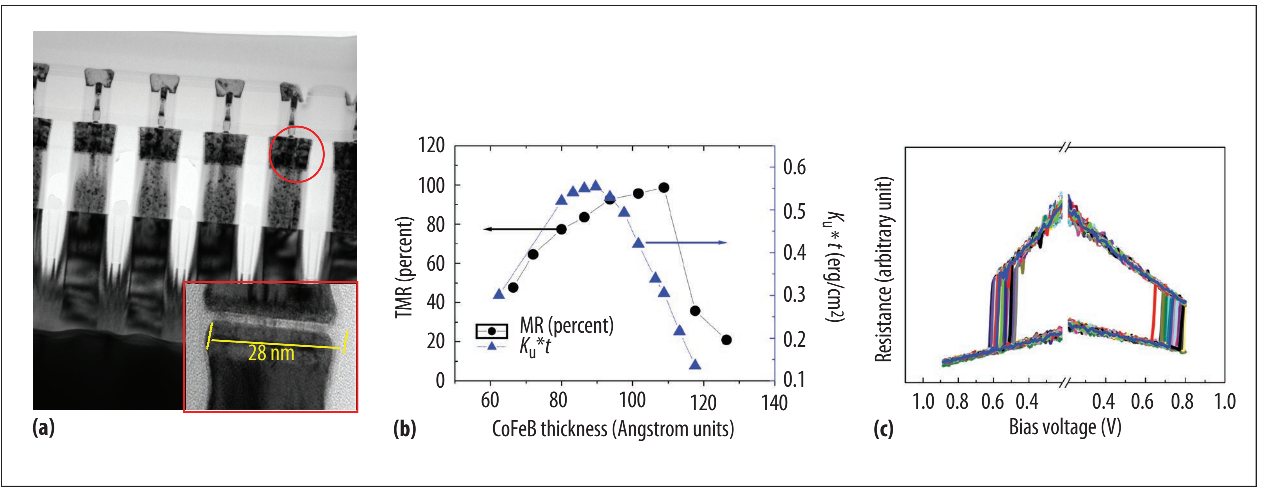


Figure 5. Magnetic tunnel junction (MTJ) cells. (a) Cross-sectional TEM image of a 28-nm MTJ cell, (b) tunnel magnetoresistance (TMR) ratio and interfacial perpendicular magnetic anisotropy (i-PMA) energy constants as a function of CoFeB thickness, where K_u is the anisotropy energy density and t is the thickness of the free layer, and (c) spin transfer torque (STT) switching curves at a 100-ns pulsewidth.

magnetic anisotropy (i-PMA) energy constants as a function of CoFeB free-layer thickness. Figure 5c shows the STT switching curve of the MTJ cell at a 100-ns pulsewidth.[11] With the optimal thickness, TMR is 100 percent, I_c is low, and magnetic switching is stable.

To overcome the decreased thermal stability factor in the small MTJ cell, some developers have proposed an advanced dual-interface cell stack that, as Figure 6a shows, introduces double magnesium oxide (MgO) tunneling layers.[12] Figure 6b shows comparative switching probability between the dual- and single-interface stack. Overall, the double interface produced more uniform distribution, improving the Δ value in the 20-nm MTJ cell size. Figure 6c illustrates the switching efficiency (I_{c0}/Δ) of the two cell schemes. Measurements of I_{c0} and Δ from the extrapolation of I_c as a function of pulsewidth show that the dual-interface stack's switching was significantly more efficient.[12] The improvement might be attributable to the enhanced tunneling source, which yielded a low I_c. Regardless of the cause, these improvements will make MRAM competitive with mainstream memory devices.

Issues and future trends

Even though MRAM technology has recently seen major enhancements, challenges remain for sub-20-nm MTJ cells, including a degraded TMR ratio during integration, a persistently high I_c, and an unsatisfactory Δ value. Future efforts must focus on optimizing the ferromagnetic electrode's composition, upgrading the MTJ stack scheme, and developing advanced MTJ patterning technology. With these breakthroughs, STT-MRAM could become a formidable competitor in commercial nonvolatile memory.

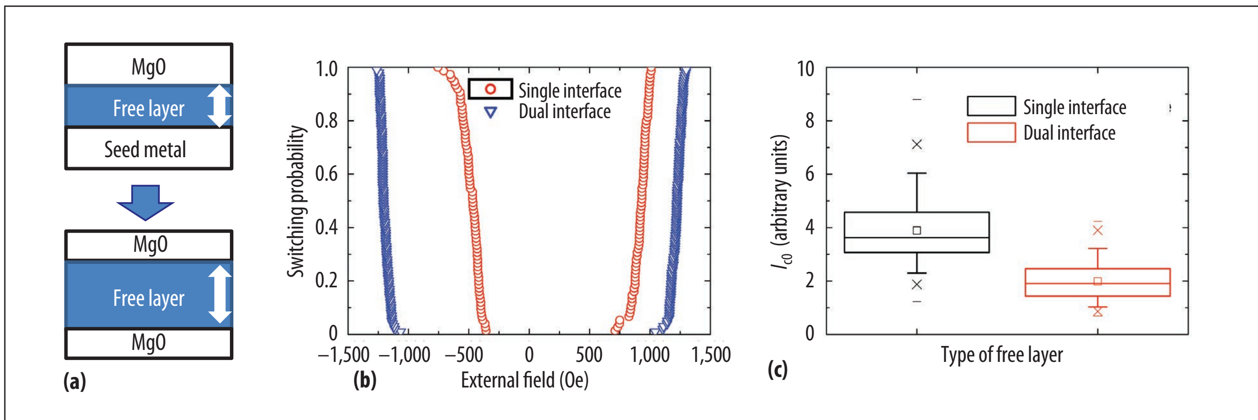


Figure 6. Single- and dual-interface stacks: (a) schematic view of a dual-interface stack, (b) comparative switching probability, and (c) switching efficiency (I_{c0}/Δ) for single- and dual-interface stacks. MgO, magnesium oxide.

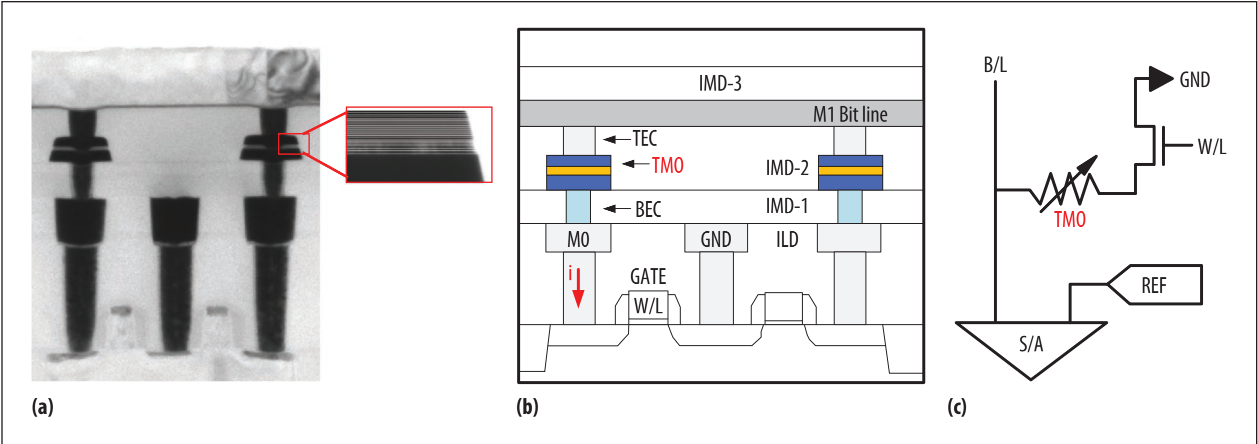


Figure 7. First transition metal oxide (TMO)-based fully integrated cell with one transistor and one ReRAM. (a) Cross-sectional TEM image, (b) schematic diagram, and (c) basic circuit diagram.

RESISTIVE RAM

ReRAM, the object of intense study from the 1960s to early 1980s, was initially attractive because it can vary its resistance according to the applied voltage and current. However, high operation voltage and current, poor endurance, and poor film-handling technique have made resistive memory less favorable than silicon-based products.

With remarkable progress in material science, however, ReRAM has started to attract the research community's interest once again as a nonvolatile, low-power, high-density, multibit operating memory technology. In 2002, ReRAM first appeared to characterize this memory type, stemming from the emphasis on "resistive" results—using perovskite oxides such as $Pr_{0.7}Ca_{0.3}MnO_3$ enabled memory switching through pulses of reasonably small voltage and current.[13]

In 2004, transition metal oxides (TMOs) replaced perovskite oxides because of the latter's critical manufacturing limitations, such as their difficult-to-control stoichiometry, which is problematic because the substrate must have a well-defined crystal structure to achieve the required film quality. Perovskite oxides also have contamination issues in a conventional CMOS process.

Figure 7 shows an early TMO design using nickel oxide with iridium electrodes. Most manufacturers were already using TMOs as high-K dielectric and capacitor materials in fabrication lines, so TMO had no contamination issues. This was a significant advantage to chipmakers because they did not have to invest any additional resources in ReRAM fabrication. Consequently, TMOs are the most popular material in ReRAM research efforts.

Also in 2002, developers began using solid electrolyte materials with silver or copper ions for another ReRAM type—conductive bridge RAM, or CBRAM. CBRAM uses diffusive metal ions to make conductive filaments, which bridge two electrodes. Oxide-based ReRAM makes oxygen ions move back and forth to generate and remove low-resistance current paths between two face-to-face electrodes.

With successive technical evolutions in scaling, megabit-density ReRAM test chips that use a one-transistor-and-one-ReRAM cell structure and TMO and CBRAM materials are emerging for embedded flash memory applications.[14]

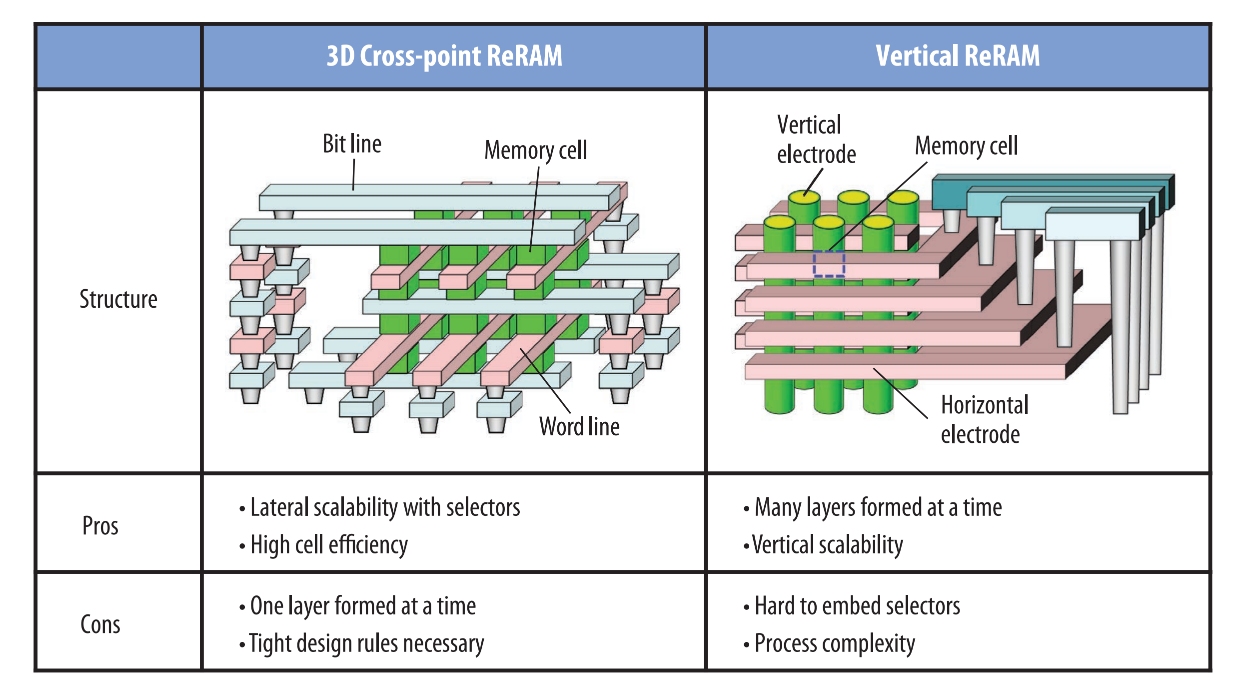
Applications

Because ReRAM operation is mainly field-induced ion migration, it has a relatively slow speed and limited endurance, but it also has relatively good data retention and a simple cell structure that is easy to manufacture. These advantages make ReRAM a strong candidate to replace flash memory technology in such applications as code storage (NOR flash) and data storage (NAND).

However, without new integration technologies, ReRAM development will face formidable obstacles. ReRAM is poised to compete with DRAM-buffered NAND flash and PRAM because, although process technology is certainly ready for code storage applications, fabrication cost remains an issue for these more mature memory technologies. The potentially low cost to fabricate ReRAM, on the other hand, makes it attractive for large-scale-integration companies. For example, generating an embedded ReRAM layer requires only two additional patterning masks after the transistor process.

Data storage like NAND flash has a slight edge over ReRAM in power consumption. Even though ReRAM has reliability advantages, low operation voltage, and random access, its power consumption is higher because it has fewer parallel working cells.

However, the most important feature for data storage application is bit density, and new ReRAM technologies look promising, as researchers investigate how 3D ReRAM



	3D Cross-point ReRAM	Vertical ReRAM
Structure	Bit line; Memory cell; Word line	Vertical electrode; Memory cell; Horizontal electrode
Pros	• Lateral scalability with selectors • High cell efficiency	• Many layers formed at a time • Vertical scalability
Cons	• One layer formed at a time • Tight design rules necessary	• Hard to embed selectors • Process complexity

Figure 8. Pros and cons of two resistive RAM (ReRAM) cell structures. 3D cross-point ReRAM literally stacks cross-point ReRAM arrays one after another, while vertical ReRAM forms all the ReRAM cells at one time after making the mold and holes.

architecture can achieve a higher bit density than planar structures. Figure 8 shows a traditional 3D cross-point array and an innovative vertical ReRAM (VReRAM) structure.

Each of these 3D ReRAM cell structures has its pros and cons. One advantage of 3D cross-point ReRAM is the ability to insert two-terminal selection devices such as diodes to reduce leakage currents without compromising lateral scalability. Positioning peripheral circuits under the cell stacks can further improve cell efficiency. On the downside, fabrication cost is high because the same process must be repeated as often as the number of stacks. Researchers have proposed programmable switch and neuromorphic applications, but these are still nascent.

Future trends

In the past decade, researchers have studied ReRAM extensively to confirm its production scalability and reliability. Its complex working mechanism—one of the hottest topics in material science—is so well understood that accurate model simulation can predict ReRAM properties with physical parameters.[15] However, for ReRAM to be a major memory product, researchers must find a way to better control its stochastic nature and variability. One thing is certain: ReRAM's production cost and reliability will decide its future place in the memory market.

TECHNOLOGY COMPARISON

Table 1 summarizes the key electrical properties of PRAM, MRAM, and ReRAM. MRAM can work as nonvolatile working memory if it solves scalability issues from patterning and thermal stability. PRAM's nonvolatility, random access, and relatively fast speed make it suitable as code memory and buffer memory. Even though ReRAM is just taking off, it will be a powerful data storage memory technology if researchers can develop a 3D stack structure.

Although mature memory choices, such as NOR and NAND flash, appear to be dominating the market, resistance-based memory's low fabrication cost and scalability are compelling to manufacturers. Work is already addressing the practical integration solutions and identifying ways to overcome obstacles to further scaling. PRAM research should focus on atomic layer deposition technology, creating a metal shield scheme to control thermal disturbance, and exploring multilevel GST. MRAM research should focus on achieving short-free etching and finding the optimal ferromagnetic electrode composition. ReRAM research must investigate 3D stacks.

Unseating mainstream memory devices that are entrenched in their application niches will require effort. For that reason, research should also focus on finding new applications that fully exploit resistance-based memory's

Table 1. Characteristics of major existing and emerging memory devices.

	DRAM	NAND	MRAM	PRAM	ReRAM
Cell size	~8F^2	~5F^2 (SLC)	~8F^2	~4F^2	~8F^2 (transistor) <1 F^2 (stack)
Density	xGigabit	xxGigabit	xxMegabit	xGigabit	--
Latency	~50 ns	20 to ~200 µs	~50 ns	~200 ns	~50 ns to ~1 µs
Bandwidth	~1 GBps	~100 MBps	~1 GBps	~100 MB/s	~100 MBps
Volatility	Volatile	Nonvolatile	Nonvolatile	Nonvolatile	Nonvolatile
Endurance	>10^{15}	>10^5	>10^{15}	>10^6 to $10^{\sim 12}$	>10^5
Retention	>64 ms	>10 years	>10 years	>10 years	>10 years
Application	Working memory	Data storage memory	Working nonvolatile RAM	Code memory and buffer memory	Data storage memory

unique properties. When solutions and novel applications emerge, at least one of these three memory technologies will begin to expand its market status.

References

1. S.R. Ovshinsky, "Reversible Electrical Switching Phenomena in Disordered Structures," *Physical Rev. Letters*, vol. 21, 1968, pp. 1450-1453.
2. S. Lai, "Current Status of the Phase Change Memory and Its Future," *IEEE Int'l Electron Devices Meeting Tech. Digest*, 2003, pp. 255-258.
3. M.J. Kang et al., "PRAM Cell Technology and Characterization in 20-nm Node Size," *IEEE Int'l Electron Devices Meeting Tech. Digest*, 2011, pp. 39-42.
4. Y.-J. Song et al., "Highly Reliable 256-Mb PRAM with Advanced Ring Contact Technology and Novel Encapsulating Technology," *VLSI Tech. Digest*, 2006, pp. 118-119.
5. H. Horii et al., "A Novel Cell Technology Using N-doped GeSbTe Films for Phase-Change RAM," *VLSI Tech. Digest*, 2003, pp. 177-178.
6. W.H. Butler et al., "Spin-Dependent Tunneling Conductance of Fe|MgO|Fe Sandwiches," *Physical Rev. B*, vol. 63, no. 5, 2001, pp. 054416-1 to 054416-12.
7. S.S. Parkin et al., "Giant Tunneling Magnetoresistance at Room Temperature with MgO (100) Tunnel Barriers," *Nature Materials,* vol. 3, 2004, pp. 862-867.
8. S. Yuasa et al., "Giant Room-Temperature Magnetoresistance in Single-Crystal Fe/MgO/Fe Magnetic Tunnel Junctions," *Nature Materials*, vol. 3, 2004, pp. 868-871.
9. M. Gajek et al., "Spin Torque Switching of 20-nm Magnetic Tunnel Junctions with Perpendicular Anisotropy," *Applied Physics Letters*, vol. 100, no. 13, 2012, pp. 132408-1 to 132408-3.
10. D.C. Worledge et al., "Spin Torque Switching of Perpendicular Ta|CoFeB|MgO-based Magnetic Tunnel Junctions," *Applied Physics Letters*, vol. 98, no. 2, 2011, pp. 022501-1 to 0220501-3.
11. W. Kim et al., "Extended Scalability of Perpendicular STT-MRAM towards Sub-20-nm MTJ Node," *IEEE Int'l Electron Devices Meeting Tech. Digest*, 2011, pp. 531-534.
12. J.H. Park et al., "Enhancement of Data Retention and Write Current Scaling for Sub-20-nm STT-MRAM by Utilizing Dual Interfaces for Perpendicular Magnetic Anisotropy," *VLSI Tech. Digest*, 2012, pp. 57-58.
13. W.W. Zhuang et al., "Novel Colossal Magnetoresistive Thin Film Nonvolatile Resistance Random Access Memory (RRAM)," *IEEE Int'l Electron Devices Meeting Tech. Digest*, 2002, pp. 193-196.
14. I.-G. Baek et al., "Highly Scalable Nonvolatile Resistive Memory Using Simple Binary Oxide Driven by Asymmetric Unipolar Voltage Pulses," *IEEE Int'l Electron Devices Meeting Tech. Digest*, 2004, pp. 587-590.
15. R. Degraeve et al., "Dynamic 'Hour Glass' Model for SET and RESET in HfO2 RRAM," *VLSI Tech. Digest*, 2012, pp. 75-77.

***Yoon-Jong Song** is a principal engineer at Samsung Electronics. His research interests include high-density PRAM development and high-density MRAM. Song received a PhD in materials engineering and science from Virginia Tech. Contact him at yoonjong.song@samsung.com.*

***Gitae Jeong** is vice president of research and development of new memory devices at Samsung Electronics. His research interests include MRAM and optical interconnection development. Jeong received a PhD in physics from Seoul National University. Contact him at gjeong@samsung.com.*

***In-Gyu Baek** is a principal engineer at Samsung Electronics. His research interests include MRAM and ReRAM development, particularly transitional metal-oxide-based resistive memories. Baek received a PhD in physics from Seoul National University. Contact him at ig.baek@samsung.com.*

***Jungdal Choi** is vice president of R&D for ReRAM, Ge, and III-V logic technologies at Samsung Electronics. His research interests also include process integration of high-density planar and 3D NAND flash memory types from 270-nm to 1x-nm design rules. Choi received a BE in electronics engineering from Kyungbuk National University. Contact him at jdchoi@samsung.com.*

Selected CS articles and columns are available for free at http://ComputingNow.computer.org.