Overview and outlook of emerging non-volatile memories

Mengwei Si, Huai-Yu Cheng, Takashi Ando, Guohan Hu, and Peide D. Ye*

Memory technologies with higher density, higher bandwidth, lower power consumption, higher speed, and lower cost are in high demand in the current big data era. In this paper, recent progress of emerging non-volatile memories is reviewed. The current status, challenges, and opportunities of emerging non-volatile memories, such as phase-change memory, resistive random-access memory, ferroelectric field-effect transistors, and magnetic random-access memory, are discussed toward storage-class memory, embedded non-volatile memories, and near/in-memory computing applications.

Introduction

In a classical von-Neumann architecture, the memory hierarchy can be divided into three levels with different response time. The first level is the registers and caches within the central processing unit (CPU), which is realized by static random-access memory (SRAM) technology made of logic transistors with a typical six-transistor (6 T) configuration. SRAM is the fastest memory with a typical read/write time of less than 1 ns and unlimited endurance, but has low area density due to the 6 T cell configuration. The second level is main memory utilizing dynamic random-access memory (DRAM), where the DRAM cell has a 1-transistor 1-capacitor (1T1C) structure. DRAM has a much higher area density than SRAM because of the 1T1C cell structure but needs periodic refreshing because of the charge loss in the capacitor. Both SRAM and DRAM are volatile memories where the information will be lost once the power is off. The third level of the memory hierarchy is storage including solid state drive (SSD) and hard disk drive (HDD). Therefore, non-volatility is required to avoid data loss when the power is off. The 3D NAND flash memory technology is the current dominating non-volatile memory technology for storage applications, because of its high density and low cost.

The performance of a computer system with von-Neumann architecture is limited more and more by the data transfer process between the CPU and the memory, known as the memory wall. The data transfer process is mainly limited by three factors:[1–3] (1) high energy consumption, where over 50% of power consumption is on data movement; (2) low bandwidth, where the bandwidth inside the memory chip is 100 times greater than the bus between CPU and memory unit such as DRAM and SSD; (3) high latency, where memory access to storage in the SSD or HDD is much slower than that of SRAM or DRAM. There are a few promising approaches to overcome this data transfer bottleneck. One approach is to develop new embedded non-volatile memory (eNVM) technologies to replace DRAM and SRAM for certain applications. The leading emerging eNVM technologies include spin-transfer torque magnetic random-access memory (STT-MRAM), phase change random-access memory (PCRAM or PCM), resistive random-access memory (ReRAM), and ferroelectric field-effect transistors (Fe-FETs). Another approach is to introduce the storage-class memory (SCM) as a main working memory solution. The SCM is expected to have similar capacity of a storage medium, but can be used as a non-volatile DRAM to address the performance gap between DRAM with fast access

Mengwei Si, Elmore Family School of Electrical and Computer Engineering, Purdue University, USA; mengwei.si@sjtu.edu.cn.
Huai-Yu Cheng, Macronix International Co., Ltd., Emerging Central Lab., Taiwan, ROC; hymcheng@us.ibm.com or hycheng@mxic.com.tw
Takashi Ando, IBM T. J. Watson Research Center, USA; andot@us.ibm.com
Guohan Hu, IBM T. J. Watson Research Center, USA; hug@us.ibm.com
Peide D. Ye, Elmore Family School of Electrical and Computer Engineering, Purdue University, USA; yep@purdue.edu
*Corresponding author
doi:10.1557/s43577-021-00204-2

© The Author(s), under exclusive license to the Materials Research Society 2021

time and high-density SSD with high latencies. Computing efficiency and power consumption can be greatly improved if data can be contained in the main working memory without frequently fetching it from storage such as SSD or HDD. The third approach is to develop new computing architectures near or in the memory, such as neuromorphic computing, to avoid or minimize the frequent memory access. Emerging non-volatile memories such ReRAM, PCM, and Fe-FETs are good candidates for neuromorphic computing applications because of their small cell size and potential to store multiple synaptic weights.

In this work, emerging non-volatile memory technologies including PCM, ReRAM, Fe-FET, and MRAM are reviewed for their potentials to overcome the memory wall challenge in a modern computing system. The four different emerging non-volatile memory technologies are reviewed in the following order: first, PCM technology is discussed with a focus on PCM devices and selectors for 3D cross-point memory application as storage-class memory; second, the principle, limitations, and potential applications of ReRAM technology is reviewed for eNVM and in-memory computing; third, Fe-FET as a 1 T non-volatile memory technology is reviewed with emphasis on the critical role of the ferroelectric (FE)/dielectric (DE)/semiconductor stack and its potential for eNVM, 3D NAND, neuromorphic computing, and edge intelligence applications; lastly, the current status and challenges/opportunities for MRAM technology is reviewed as eNVM and a potential candidate to replace SRAM for last level cache applications.

Phase-change memory

PCM is enabled by the phase transition of a chalcogenide material (containing one or more elements of group VI, such as S, Se or Te, or so-called phase-change materials) between two different states: an amorphous state that shows very low electrical conductivity; and a polycrystalline state that exhibits high conductivity. The phase transition is triggered by electric pulses (or heat surges) (**Figure 1**).[4] The "SET" operation switches the memory cell to the polycrystalline state: a moderate magnitude voltage pulse is applied to the material within a cell to heat it above its crystallization temperature. Conversely, "RESET" operation switches the cell back to its amorphous state: a large and short electric pulse melts the polycrystalline phase into a transient liquid state, which is then rapidly quenched by the cold surroundings into an amorphous state. The "Read" operation can be performed nondestructively (read out) by measuring the memory cell resistance. The difference in resistance between the two states is known to be three to six orders of magnitude, which provides a sufficient margin for sensing.

Due to the increase of the demand for handling a large amount of data, modern computer systems need more and more memory (DRAM) to handle processing. Computing efficiency and power consumption can greatly improve if the data can be maintained in the main working memory without frequently fetching it from the storage media (SSD or HDD). The concept of SCM[5] is to provide a main working memory solution, which has similar capacity of a storage media, but can be treated as a nonvolatile DRAM. Therefore, the SCM can narrow the performance gap between fast latencies DRAM and high-density storage media (like NAND Flash). To achieve the goal of SCM, the first challenge is the density/capacity. Conventional, planar PCM technologies consists of a single layer of memory devices and the access devices are built directly on the single-crystalline silicon substrate. For the past 10 years, the PCM density continuously improved thanks to device-to-device pitch scaling and scaling of PCM device size. This also leads to a reduction in required programming power (RESET power). 1 Gbit 45 nm node (effective cell size of 5.5 F^2, where F is the minimum feature size) PCM chip[6] was demonstrated using a vertical pnp-BJT as an access device. The BJT could deliver 300 μA at 2.0 V, with a base-emitter leakage current in reverse bias of less than 0.1 pA/bit at 3 V. A polysilicon diode on metal lines with the integration on top of peripheral circuit but below M1 metallization (the first metal layer) was demonstrated a 1 Gb array with 4 F^2 cell at the 84 nm pitch (in this case $F = 42$ nm).[7] In their work the polysilicon diode can successfully switch a PCM cell with a RESET current of approximately 220 μA (~12 MA/cm^2) and achieves a SET/RESET endurance above 10^8 cycles. With aggressive scaling of the PCM device cell, an atomic-layer-deposited (ALD) phase-change material integrated with the similar polysilicon diode as an access device enabled an 8 Gbit PCM chip in 20 nm node with a 4 F^2 cell size,[8,9] which is highest density demonstration for PCM in conventional planar PCM technologies.

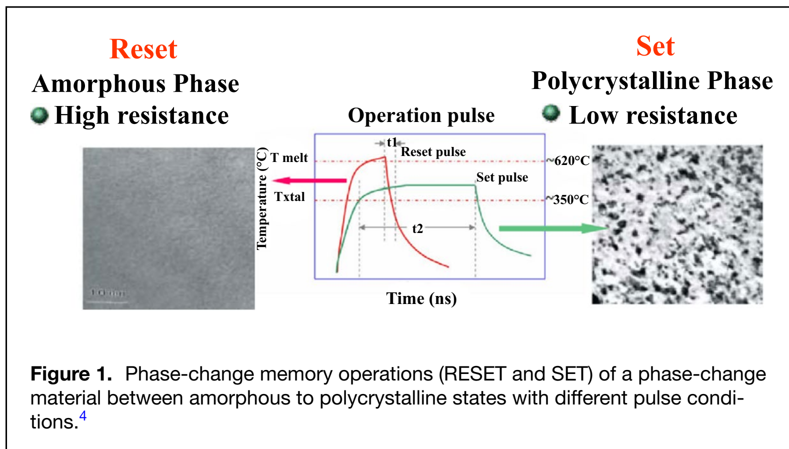


Figure 1. Phase-change memory operations (RESET and SET) of a phase-change material between amorphous to polycrystalline states with different pulse conditions.[4]

However, in order to further boost the memory density, the memory array layers need to be stacked up to form a 3D memory structure. In References 7, 8, and 9, no

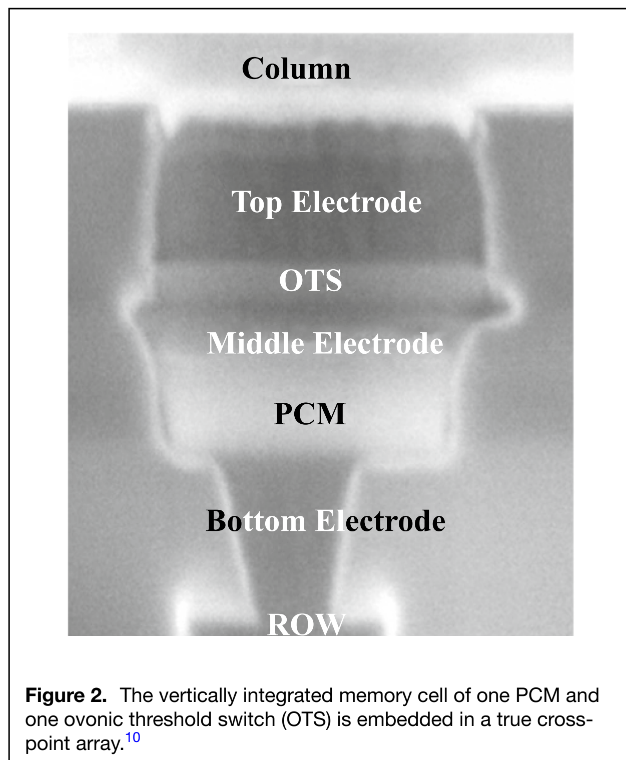


Figure 2. The vertically integrated memory cell of one PCM and one ovonic threshold switch (OTS) is embedded in a true cross-point array.[10]

specific details are given about the polysilicon deposition step and it is difficult to estimate whether these integrations schemes would be 3D stackable in a standard Cu back-end-of-line (BEOL) integrated wafer to boost the density. Generally speaking, 3D stacking on BEOL requires processing temperature below 400℃, and at this low temperature it is difficult to achieve polysilicon with fairly large grain size, which is needed to guarantee access devices with high enough ON current and low OFF leakage. Even though high performing polysilicon access devices in large arrays have been demonstrated, it is still missing a demonstration of a multi-layer stacked cross-point PCM array based on polysilicon technology, which is needed to enable high-density and cost-effective production.

Stackable selector devices are required in order to further increase the density of PCM technology. Kau et al. in Reference 10 demonstrated the feasibility of a cross-point array where an ovonic threshold switch (OTS) device is placed in series with phase-change material in a one selector/one resistor (1S1R) configuration (**Figure 2**). One of the advantages of OTS access devices over the polysilicon diode comes from the decreased number of steps involved in its integration. It only takes one additional PVD deposition to add an OTS access device to the stack, which can be done in the same cluster tool as metallization and phase-change layers for optimal throughput. In addition, such stackable structures can be packed in cross-point arrays in only 4 F^2 cell size boosting the density of the PCM technology. This OTS material is based on one kind of chalcogenide alloy glasses which was first demonstrated by Ovshinsky in the 1960s[11] where an amorphous semiconductor $Te_{48}As_{30}Si_{12}Ge_{10}$ chalcogenides exhibited a rapid and nondestructive reversible transition between a highly resistive and conductive state effected by an electric field (called threshold switching). It can be heated to the molten state and cooled slowly without devitrification.[12]

The access devices and memory devices are both based on chalcogenides materials and their different switching characteristics (either threshold or memory) depends on the chalcogenides composition and can be tuned by appropriate material designs. OTS needs to satisfy few criteria to be successfully employed as access device: stay in amorphous state after fabrication and during high-current operation; provide high enough current to be able to successfully RESET PCM element; and critically important is to have a low enough leakage current to achieve a successful program and read procedure avoiding sneak paths issue and disturbance of half-selected cells during cross-point arrays operation. Detailed cross-point operations and further considerations are described in Reference 13.

As mentioned earlier, OTS material based on TeAsGeSi system were demonstrated for a selector device in 1968[11] and have been the primarily studied materials. However, insufficient cycling endurance and low thermal stability remains a key hurdle that inhibits these materials to be used in a large cross-point array. N_2 plasma treatment on TeAsGeSi material has been demonstrated to improve the thermal stability by providing a barrier for Te loss during the high temperature annealing process,[14] as a result the thermal stability and threshold voltage (V_{th}) distribution is retained after many cycles. In addition, different Ge, Si, and As concentrations in the TeAsGeSi system[15] are tuned and optimal composition is proposed to reach endurance performance to 10^{11} cycles. However, the high leakage current (I_{OFF}) and high V_{th} drift are still the main concerns for TeAsGeSi-based materials.

Further improvement of TeAsGeSi OTS material by incorporation of Se and another dopant (<5 at.%) has been successfully demonstrated. Although such chalcogenide OTS material is relatively complex, with special material design and process control, an excellent selector can be achieved.[16] In addition, carefully controlling the reactive-ion etching (RIE) process using an extra buffer layer as a stop layer when integrating OTS with PCM makes it possible to prevent process-induced modification of both OTS and PCM materials. 1S1R OTS with the above-mentioned material has been successfully integrated with PCM (**Figure 3**a).[17] The ON-current of this OTS can reach 850 µA, which means a high current density of 8.9 MA/cm^2. A RESET speed of 10 ns, read endurance of 10^{11} cycles (Figure 3b, OTS is turn ON) and program endurance of 10^9 cycles were demonstrated. The device shows nearly identical *I*–*V* characteristics before and after anneals at 400℃/30 min, thus demonstrating compatibility with further BEOL processing.

Se-based chalcogenides have also been proposed as selectors.[18] It was originally designed by replacing the main chalcogenide element of the selector (Te) with Se and simplifying

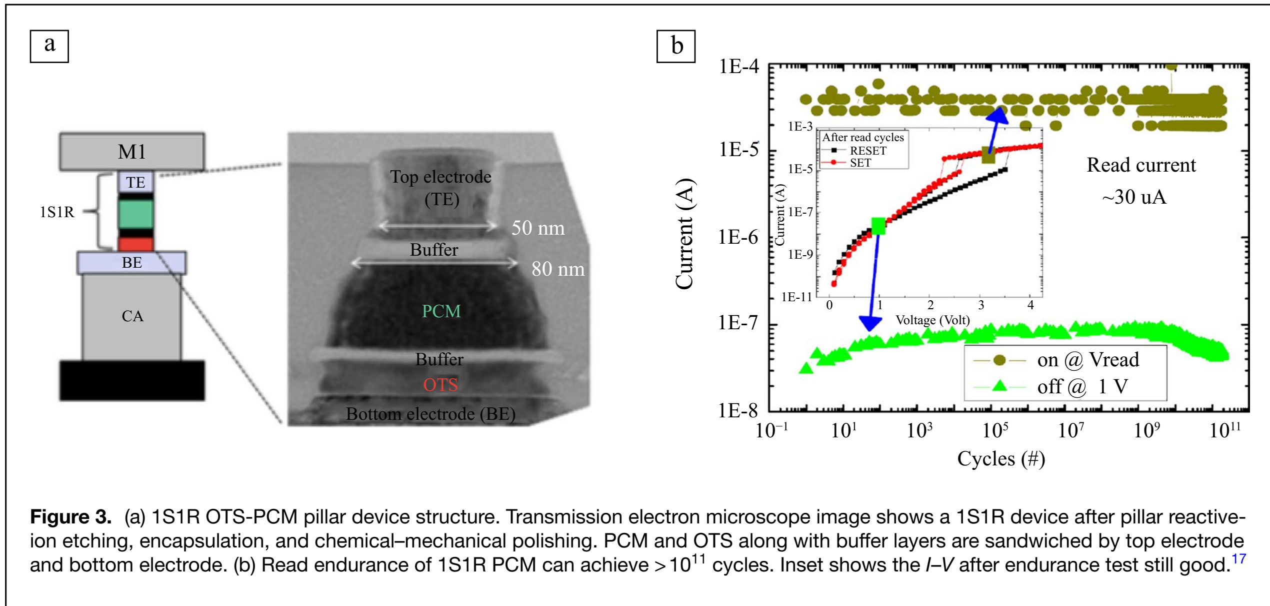


Figure 3. (a) 1S1R OTS-PCM pillar device structure. Transmission electron microscope image shows a 1S1R device after pillar reactive-ion etching, encapsulation, and chemical–mechanical polishing. PCM and OTS along with buffer layers are sandwiched by top electrode and bottom electrode. (b) Read endurance of 1S1R PCM can achieve $>10^{11}$ cycles. Inset shows the *I–V* after endurance test still good.[17]

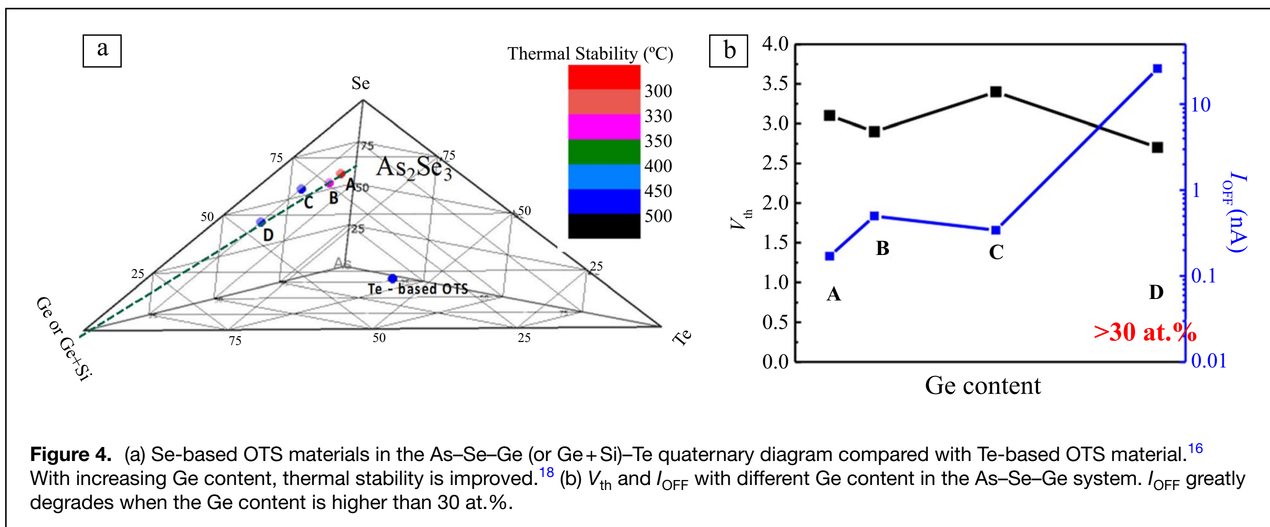


Figure 4. (a) Se-based OTS materials in the As–Se–Ge (or Ge + Si)–Te quaternary diagram compared with Te-based OTS material.[16] With increasing Ge content, thermal stability is improved.[18] (b) V_{th} and I_{OFF} with different Ge content in the As–Se–Ge system. I_{OFF} greatly degrades when the Ge content is higher than 30 at.%.

the system to a ternary compound (As–Se–Ge). With increasing the Ge content for the Se-based selector along the As_2Se_3-Ge tie-line, the thermal stability can be significantly increased (**Figure 4**a); however, I_{OFF} degrades significantly when the Ge content of chalcogenides along the As_2Se_3-Ge tie-line is higher than 30 at.% (Figure 4b). Low Ge content selector exhibits superior cycling endurance (6.9×10^{11}), which is the best among OTS chalcogenides along this tie-line.[18] Cycling endurance degrades with further increase of Ge content. Although higher Ge concentration guarantees better thermal stability (Figure 4a), it has the drawback of, for this particular system, yielding a device with unstable cycling endurance and high I_{OFF}; on the other hand, low Ge material exhibits extraordinary cycling endurance and low I_{OFF} but poor thermal stability. The trade-off between thermal stability, I_{OFF}, and cycling endurance is critical when choosing the optimum composition with respect to Ge content. Si effectively increases both the thermal stability and the optical bandgap,[15] indicating that Si incorporation into the AsSeGe system is expected to achieve higher thermal stability without significantly sacrificing I_{OFF}.[19]

From an integration perspective a thinner OTS is preferred. V_{th} decreases with decreasing OTS thickness for AsSeGe material with and without Si incorporation (**Figure 5**a).[19] However, I_{OFF} exhibits a sharper increase with decreasing OTS thickness for the AsSeGe material. 20 nm OTS with Si incorporation shows very good electrical characteristics ($V_{th} = 3.4$ V and good $I_{OFF} = 381$ pA@2 V), making this a better candidate for small pitch high-density memory, compared to undoped AsSeGe. AsSeGeSi achieves $> 2 \times 10^{11}$ cycling endurance (Figure 5b). Additionally, all of the 66 test cells in 30-nm-thick OTS devices pass 10^9

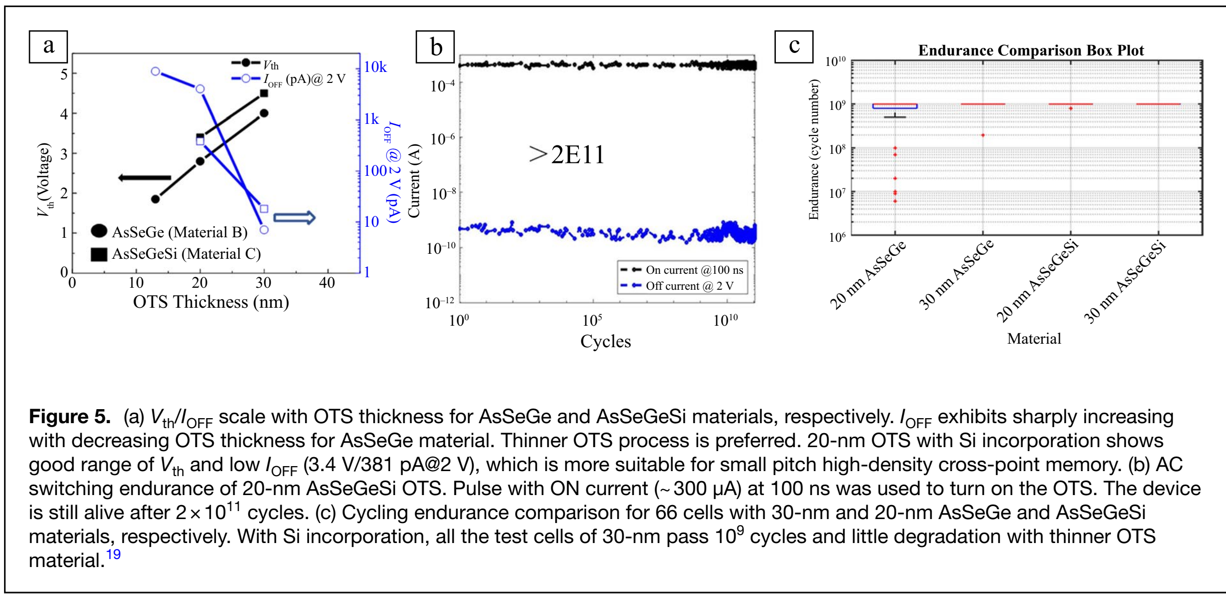


Figure 5. (a) V_{th}/I_{OFF} scale with OTS thickness for AsSeGe and AsSeGeSi materials, respectively. I_{OFF} exhibits sharply increasing with decreasing OTS thickness for AsSeGe material. Thinner OTS process is preferred. 20-nm OTS with Si incorporation shows good range of V_{th} and low I_{OFF} (3.4 V/381 pA@2 V), which is more suitable for small pitch high-density cross-point memory. (b) AC switching endurance of 20-nm AsSeGeSi OTS. Pulse with ON current (~300 μA) at 100 ns was used to turn on the OTS. The device is still alive after 2×10^{11} cycles. (c) Cycling endurance comparison for 66 cells with 30-nm and 20-nm AsSeGe and AsSeGeSi materials, respectively. With Si incorporation, all the test cells of 30-nm pass 10^9 cycles and little degradation with thinner OTS material.[19]

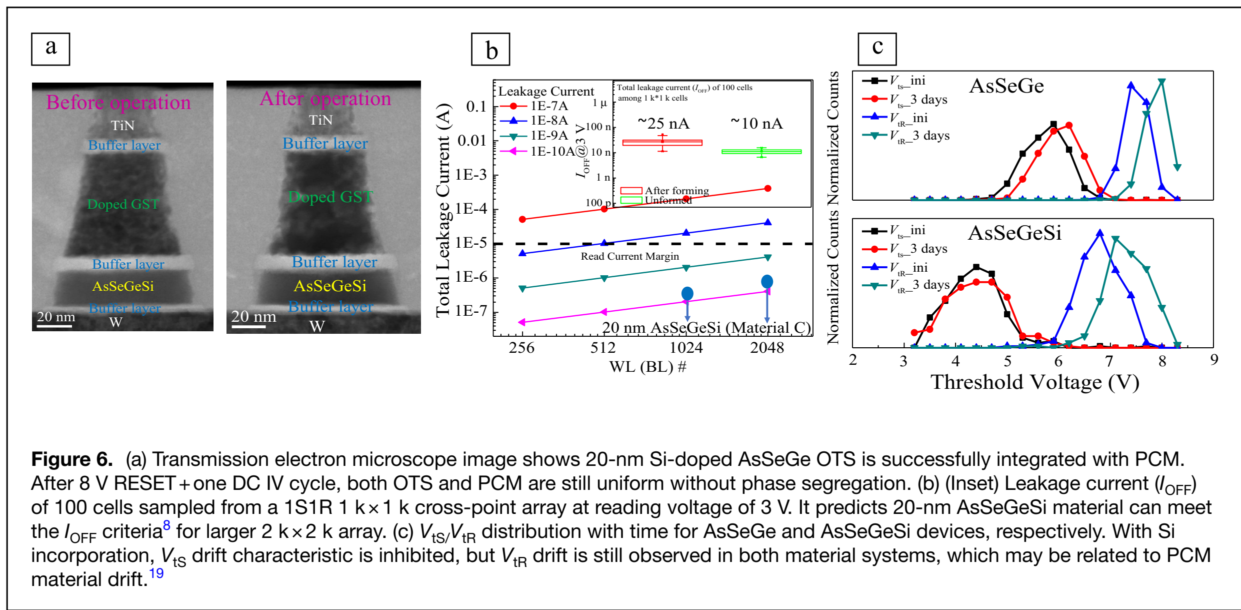


Figure 6. (a) Transmission electron microscope image shows 20-nm Si-doped AsSeGe OTS is successfully integrated with PCM. After 8 V RESET + one DC IV cycle, both OTS and PCM are still uniform without phase segregation. (b) (Inset) Leakage current (I_{OFF}) of 100 cells sampled from a 1S1R 1 k × 1 k cross-point array at reading voltage of 3 V. It predicts 20-nm AsSeGeSi material can meet the I_{OFF} criteria[8] for larger 2 k × 2 k array. (c) V_{tS}/V_{tR} distribution with time for AsSeGe and AsSeGeSi devices, respectively. With Si incorporation, V_{tS} drift characteristic is inhibited, but V_{tR} drift is still observed in both material systems, which may be related to PCM material drift.[19]

cycles (Figure 5c). Without Si, the cycling endurance greatly degrades for thinner OTS thicknesses.

AsSeGeSi OTS material is successfully integrated with doped GST phase-change material in 1S1R (**Figure 6**a) cross-point arrays. Figure 6b shows low I_{OFF} (25 nA @ 3 V) from a 100 cells sampled from 1 k by 1 k 1S1R cross-point arrays, which proves that this material can meet the I_{OFF} criteria for even larger arrays (2 k by 2 k). Figure 6c demonstrates a large memory window (~2 V main distribution window) between SET and RESET V_{th} distribution (V_{tS} and V_{tR}) in the cross-point ADM array before drift. After 3 days, 0.3 V V_{tS} drift is observed for AsSeGe, but it is inhibited for system with Si incorporation (non-detectable V_{tS} drift), which is very promising for cross-point memory technology.

With continuous improvement on OTS materials, as well as careful process consideration during integration, 128 Gb 2-layer 20 nm half-pitch cross-point technology[20] and a 4-layer cross-point technology have been realized, greatly boosting the memory density to 256 Gb (**Figure 7**).[21]

High-density commercial products show the feasibility of 3D cross-point PCM technology.[21] Continuous efforts and research on both OTS and PCM materials improvement could eventually decouple the unpleasant trade-off requirements for the access and memory device and circuit design optimization will decrease latencies and increase bandwidth, which enables SCM solution in new system architectures.

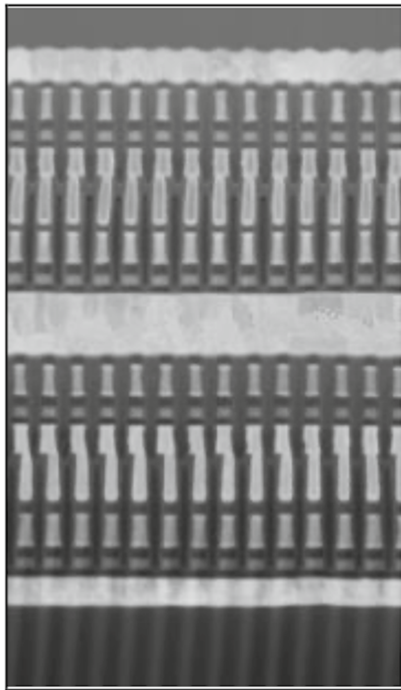


Figure 7. Four layers of 3D cross-point technology based on OTS and PCM layers in 20-nm half-pitch for 256 Gb memory density.[21]

Resistive random-access memory

It has been known for more than 50 years that a metal–insulator-metal (MIM) structure can switch between an insulating state and a conducting state depending on choice of materials.[22] Although a simple MIM structure functioning as a memory was an attractive finding, no sophisticated process technology that can precisely control oxide thickness in the nanometer scale has been available in the early days. Therefore, resistive oxide devices have remained as a research topic until recently. As film thicknesses in the state-of-the-art semiconductor devices scale below 10 nm, many film deposition techniques, such as ALD, emerge to support manufacturing needs. As a result, oxide-based resistive memory devices attracted a renewed interest. There are multiple device options in the community, and they can be categorized into oxide ReRAM[23] and conductive bridging random-access memory (CBRAM),[24] depending on the elements that contribute to the current conduction. Oxide ReRAM can be further categorized into filamentary and non-filamentary types. In order to initiate filamentary oxide ReRAM devices, an operation called electro forming is required. This process is analogous to soft breakdown of oxides and can be predicted accurately with the statistical model of oxide breakdown.[25] The switching mechanisms of filamentary oxide ReRAM and CBRAM are schematically summarized in **Figure 8**. After an electro forming process, a current conducting filament is formed within the metal oxide layer. By applying electrical bias on the electrode, the configuration of the filament is changed, resulting in resistive switching between low resistance state (LRS) and high resistance state (HRS). The operation to switch from HRS to LRS is called set and the opposite switching is called reset.

One of the advantages of filamentary ReRAM is that relatively simple metal oxides which have been already used in semiconductor manufacturing, such as HfO_x and TaO_x, can be used, making it highly compatible with CMOS technology. It is a promising approach for future NVM device applications due to its fast speed (~ns), scalability to the nanometer regime, and ultra-low power consumption (~nW). On the other hand, one of the big challenges is the stochastic nature of the filament, which results in device-to-device variability and abrupt switching behaviors. In a non-filamentary ReRAM, defect migration takes place over the entire device area, at an interface between two materials such as an oxide and metal.[26] This modulates the Schottky barrier of the interface and enables resistive switching. While this type of switching avoids stochasticity due to filaments, it tends to show slower switching speed and degraded retention. CBRAM has a switching mechanism similar to that of the filamentary oxide ReRAM, but it utilizes metal cations, such as Ag and Cu, as building blocks of current conducting filaments. Since these materials are used for metallization, this option is also highly compatible with CMOS technology.

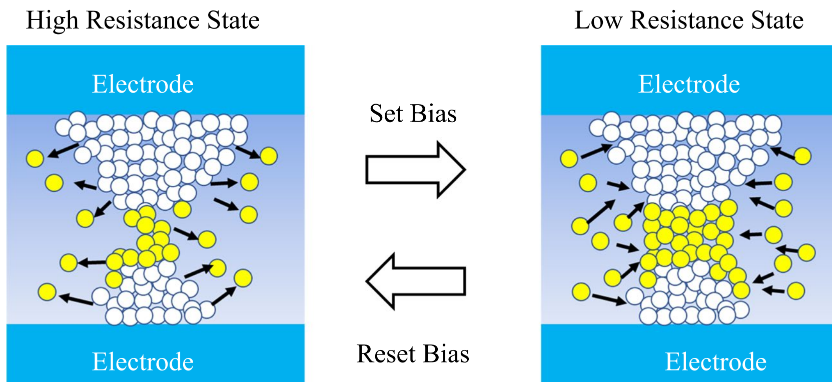
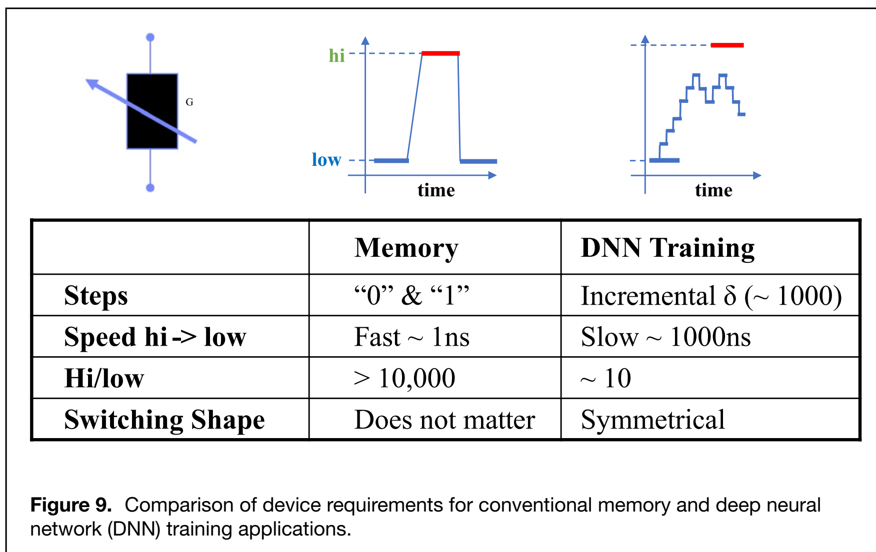


Figure 8. Schematic illustration of switching mechanism of filamentary oxide resistive random-access memory (ReRAM) and conductive bridging random-access memory (CBRAM). The building blocks of the filament are oxygen vacancies for oxide ReRAM and metal cations for CBRAM.

In order to justify ReRAM as a standalone non-volatile memory, the memory density needs to be competitive against other candidates, such as NAND flash memory. Along this thread, a 3-dimensional



	Memory	DNN Training
Steps	"0" & "1"	Incremental δ (~ 1000)
Speed hi -> low	Fast ~ 1ns	Slow ~ 1000ns
Hi/low	> 10,000	~ 10
Switching Shape	Does not matter	Symmetrical

Figure 9. Comparison of device requirements for conventional memory and deep neural network (DNN) training applications.

structure (vertical ReRAM) was proposed;[27] however competing with ever evolving NAND flash memory with density exceeding 1 Tb is not easy. On the other hand, due to the high compatibility of oxide ReRAM and CBRAM to conventional CMOS technologies, these memories showed good promise as embedded memories.[28,29] For this type of application, endurance of oxide ReRAM and CBRAM, which is typically around 10^5, is the limiting factor and it requires further improvement before it becomes a more versatile embedded memory option.

New emerging applications of ReRAM are in-memory computing or computational memory, which exploits the physical properties of non-volatile memory devices for both storing and processing information. One such paradigm is hardware acceleration of deep neural network (DNN) training via the use of dense crossbar arrays of ReRAM to perform locally analog computation at the location of the data. The comparison of device requirements for conventional memory and DNN training is summarized in **Figure 9**. One of the key requirements to implement a conventional DNN training algorithm is that resistive devices must change conductance in a symmetrical fashion when subjected to positive or negative pulse stimuli.[30] Most ReRAM devices exhibit highly asymmetrical conductance change, which results in significant errors in weight updates. In addition, when incremental weight updates are performed for ReRAM devices, the magnitude of conductance change approaches the level of inherent randomness of the materials, manifesting as significant noise components.[31] Exploration of ideal symmetrically switching devices requires deviation from well-established memory technologies and falls into a long-term research agenda. Therefore, co-optimization of device and algorithm may be a more feasible approach to utilize conventional NVM devices, such as ReRAM. In this context, a modified DNN training algorithm has been recently proposed to significantly relax the requirement on the switching symmetry and to improve tolerance for device programming noise.[32] Although it is difficult to achieve perfectly symmetric switching, it was found that ReRAM devices have at least a point where the device responses to upward/downward pulses are identical, which is defined as a "symmetry point (SP)."[33] The new algorithm adopts two separate systems to leverage the presence of SP: A and C. A acts as a gradient accumulation layer, whereas C acts as a weight layer. The accumulated gradient is transferred to C with a pre-defined rate. Once the desirable weight is reached for C, the accumulated gradient fluctuates around zero and naturally brings A to the SP. Hence, at this steady state of the coupled system, learning converges, and C stores learned parameters of the model. This kind of algorithmic innovation opens pathways toward large-scale in-memory compute applications utilizing existing ReRAM technologies.

Ferroelectric field-effect transistors

Ferroelectric material exhibits a spontaneous polarization which is reversable by external electric field. The two stable polarization states making it attractive for non-volatile memory applications. Ferroelectric random-access memories (FeRAMs) with a 1T1C configuration have been in mass production, featured with low power consumption, fast speed, high data reliability even at high temperatures. A Fe-FET is a metal–oxide–semiconductor (MOS) transistor whose gate insulator film is ferroelectric, the concept of which was introduced back in the 1950s.[34] Fe-FET as a 1 T type FeRAM are preferred against 1T1C type FeRAM because of the non-destructive readout and 1 T cell configuration for high-density integration. However, Fe-FETs made of conventional perovskite ferroelectric insulators such as $Pb(Zr,Ti)O_3$ (PZT) and $SrBi_2Ta_2O_9$ (SBT) have never been commercialized because (1) a CMOS-compatible thin-film deposition process is lacking and it is very challenging to form high-quality ferroelectric insulators and ferroelectric/semiconductor interface on top of Si substrates, resulting in the poor retention characteristics;[35,36] (2) it is difficult to scale the size of conventional perovskite ferroelectric oxides in both lateral and vertical dimensions, limiting the storage density.

The hafnium oxide (HfO_2) crystal was found to be stabilized in orthorhombic phase with ferroelectricity in 2011[37] by incorporating Si as dopants, and later with many other dopants such as Al, Zr, Y, etc.[38] Among which, Zr-doped HfO_2 or hafnium zirconium oxide (HZO) featured with the highest FE properties at 50% Zr doping ($Hf_{0.5}Zr_{0.5}O_4$) so that achieving a broader dopant range than other doped FE HfO_2 like Si-doped FE HfO_2. The discovery of FE HfO_2 immediately attracted

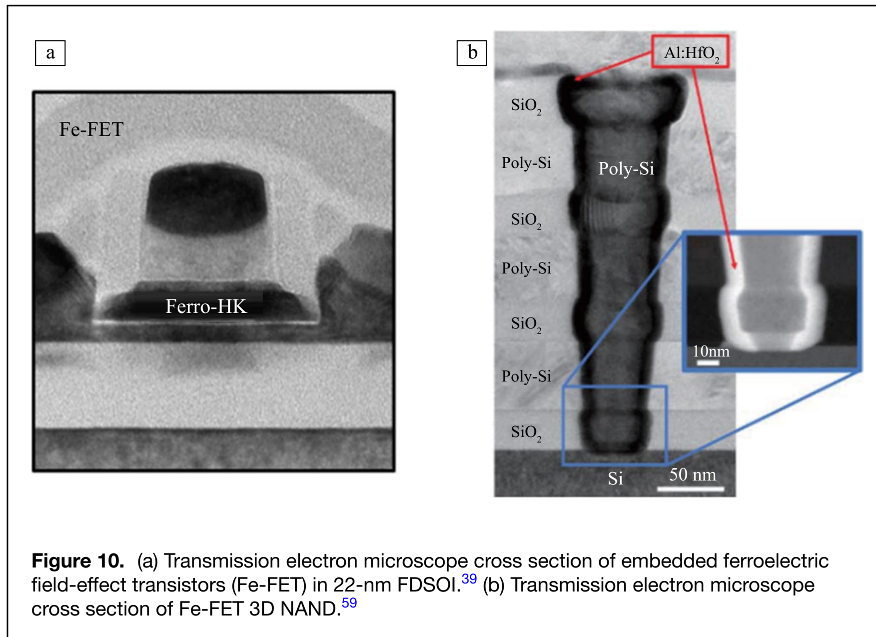


Figure 10. (a) Transmission electron microscope cross section of embedded ferroelectric field-effect transistors (Fe-FET) in 22-nm FDSOI.[39] (b) Transmission electron microscope cross section of Fe-FET 3D NAND.[59]

tremendous attention because (1) the material is fully compatible with Si CMOS technology by ALD; (2) the lateral and vertical sizes of FE HfO_2 can be scaled down to few nm and FE HfO_2-based Fe-FETs have been demonstrated in advanced technology nodes, as shown in **Figure 10**a;[39–42] (3) the high coercive field (E_C) feature of FE HfO_2 overcomes the retention bottleneck of Fe-FETs with perovskite ferroelectric oxides, with 10-year retention time demonstrated.[43,44] FE HfO_2 also exhibits fast speed with switching time less than 1 ns[45–47] and high endurance over 10^{12} cycles,[48] to be comparable with conventional perovskite ferroelectric oxides.

A Fe-FET structure is needed as non-volatile memory devices with FE HfO_2 due to the nondestructive readout. But one of the limiting factors of current Fe-FET technology is the endurance performance, for example, Si Fe-FET can typically achieve endurance on the order of 10^4–10^6 cycles,[49–51] which is many orders of magnitude lower than the endurance of a FE capacitor with metal/FE/metal (M/FE/M) structure. This large discrepancy is understood by the different structure of Fe-FET with metal/FE/DE/semiconductor structure.[52–54] A DE layer, such as SiO_2, is usually necessary to improve the oxide/semiconductor interface. The above problem comes from the high polarization charge density in FE HfO_2 with a typical FE polarization at about 10–30 $\mu C/cm^2$, while the maximum charge density supported by a dielectric material is typically a few $\mu C/cm^2$. For example, SiO_2 has a breakdown electric field of about 10 MV/cm and a dielectric constant (k) of 3.9, corresponding to a charge density of 3.5 $\mu C/cm^2$, which is far below the FE polarization density. As a result, the FE polarization switching in a Fe-FET is assisted by the charge injection to the FE/DE interface for charge balance,[46,52,54] which is verified by experiments on capacitors with FE/DE stack and Fe-FETs.[52,55] Because of the charge injection at the FE/DE interface, the memory window (MW) of a Fe-FET is determined by the strength of the DE layer (or in other words, the leakage current level at a given voltage). The MW of Fe-FET degrades during the endurance measurement because the DE layer becomes leakier due to the multiple cycles of charge injection through the DE layer. This happens before the FE layer degrades. Therefore, the endurance characteristics of a Fe-FET is determined by the DE layer instead of the FE layer. This is fundamentally different from a M/FE/M capacitor.[54]

To improve the endurance characteristics, there are two possible approaches. First, to develop a FE thin film with FE polarization density less than few $\mu C/cm^2$, so that DE layer does not operate near breakdown. The reduction of Hf composition in FE HZO can reduce the remnant polarization (P_R) of the FE layer. The P_R is also reduced rapidly once the FE HZO thickness is smaller than 4 nm.[41] Second, to develop Fe-FETs without a DE layer. The control of interface trap density (D_{it}) becomes critical for this approach because D_{it} may also screen the FE polarization which leads to the reduction of MW and endurance performance. A back-gate Fe-FET with oxide semiconductor channel[56] is preferred since no native interfacial oxide layer is needed. Recently, back-gate Fe-FETs with no DE layer with endurance $>10^8$ cycles are successfully demonstrated.[57,58]

The emerging applications of Fe-FET include eNVM, 3D NAND memory, Fe-FET neurons and synapses for neuromorphic computing applications, edge intelligence applications, etc. Fe-FET has a much lower operation voltage compared to embedded flash memory (eFlash) devices, the supply voltage of which may potentially be compatible with CMOS logic devices. Fe-FET memory with 3D NAND architecture is also being explored, as shown in Figure 10b,[59,60] which is necessary to further boost the storage density. The multi-domain nature of FE materials enables the capability to store multiple weights in a single device, which is being investigated as both synapse[61,62] and neurons[63] for neuromorphic computing applications. The non-volatile and multiple states properties of Fe-FETs also make it attractive for edge intelligence applications, where local data analysis and decision making are needed but power supply from the environments is scarce.[64]

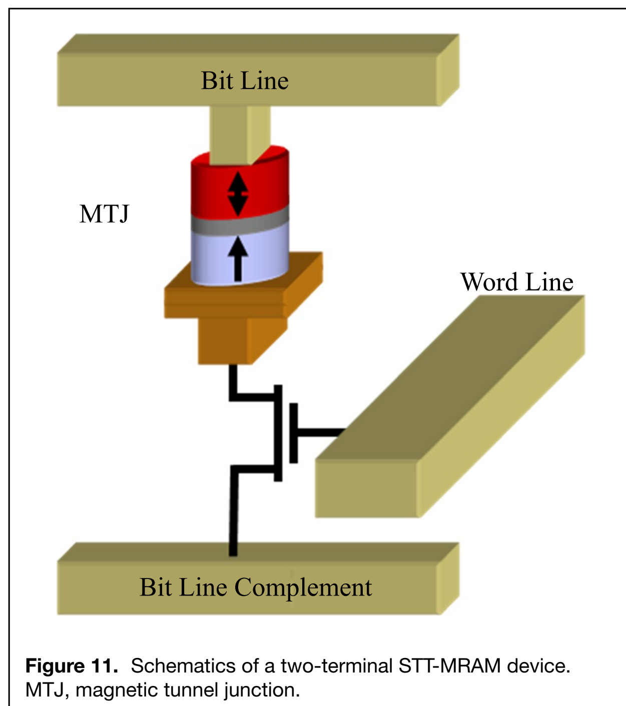


Figure 11. Schematics of a two-terminal STT-MRAM device. MTJ, magnetic tunnel junction.

Magnetic random-access memory

Spin-transfer-torque magneto-resistive random-access-memory (STT-MRAM) is a type of emerging non-volatile memory, where information is written using electron spins. A typical STT-MRAM memory cell uses a 1 transistor, 1 magnetic tunnel junction (MTJ) structure, illustrated in **Figure 11**, where both the write and read operations are achieved by passing an electrical current through the MTJ. The readability of an STT-MRAM device is determined by tunneling magneto-resistance (TMR). In an MTJ, TMR is defined as $(R_{AP}–R_P)/R_P \times 100(\%)$, where R_{AP} and R_p are the tunneling resistance of the MTJ when the two magnetic layers are in anti-parallel and parallel alignment, respectively. The larger the TMR, the bigger the difference between the two resistance states. Large TMR is desirable for fast read operation. The writing mechanism of the STT-MRAM device originates from the theoretical prediction that spin-polarized electrical current can generate a torque to switch the magnetic moment of a magnet through the transfer of angular momentum.[65–67] The first experimental demonstrations of STT writing of magnetic materials came in 1999 and 2000.[68,69] A low switching current is desirable for write operation. This section will review the major historical materials breakthroughs in the STT-MRAM field, the current status of the STT-MRAM technology, and the challenges/opportunities to advance the technology further.

The advances of STT-MRAM technology in the past 20 years have been driven by materials innovations and breakthroughs, including high TMR MTJs with MgO tunnel barriers for read operation and CoFeB-based materials with perpendicular magnetic anisotropy (PMA) for write operation.[70–80] The early generations of MTJ devices used an amorphous AlO_x tunnel barrier sandwiched between two 3d ferromagnetic electrodes. Room temperature (RT) TMR of MTJs with AlO_x tunnel barriers was first demonstrated by Miyazaki et al.[70] and Moodera et al.[71] in 1995 and was improved of up to 70–80% at RT in the following years. MTJs with an AlO_x tunnel barrier were key building blocks for earlier generations of TMR recording heads in hard disk drives as well as field-MRAM products, where the MTJs were written by magnetic field. Different from recording heads and field MRAM, STT-MRAM requires MTJs with not only high TMR, but also low resistance-area product (RA) to avoid tunnel barrier breakdown during the write operation. MTJs with AlO_x tunnel barriers, where the typical RA is from a few hundreds to thousands of Ω-μm^2, cannot fulfill either of the requirements. In 2001, Butler et al. predicted that MTJs with a (100) oriented crystalline MgO tunnel barrier combined with perfectly aligned Fe electrodes, can have a TMR >1000%, based on first-principles calculations.[72] In 2004, Yuasa et al. and Parkin et al. demonstrated ~200% TMR at room temperature in MgO-based MTJs independently.[73,74] The highest RT TMR up 600% was demonstrated in MgO-based MTJs in 2008.[75] Compared to AlO_x-based MTJs, where TMR is low (<100%) and RA is high (>100 Ω-μm^2), MgO-based MTJs enabled a combination of high TMR and low RA in the same device, which is critical to the emergence and advancement of the STT-MRAM technology. Today, the MgO-based STT-MRAM devices have a typical RT TMR ~100–200% at RA ~10–20 Ω-μm^2.

Writing of STT-MRAM devices is through passing electrical current through the MTJs. Since the early days, one of the key challenges for the STT-MRAM write operation has been to lower the write current. It was recognized early on that materials with PMA can lower the device switching current significantly compared to their counterparts with in-plane magnetic anisotropy, with the same device thermal stability.[76] The most important materials breakthrough came in 2010, when researchers at Tohoku University and IBM independently discovered the CoFeB-based materials with PMA.[77–79] When combined with the MgO tunnel barrier, CoFeB with PMA is the ideal free layer material to achieve both high TMR and low switching current in STT-MRAM devices. Researchers at IBM also developed a novel synthetic anti-ferromagnetic (SAF) reference layer structure for perpendicular MTJs (p-MTJs).[80] The reference layer is composed of three parts: (111) textured Co|Pt, or Co|Pd multilayers coupled in anti-parallel alignment through a Ru spacer, a thin Ta or W spacer as a structure transition layer, and a CoFeB|Fe|Co-based interfacial layer interfacing the MgO tunnel barrier. This SAF reference layer structure enables low offset field and high TMR of p-MTJs and is also stable under the spin-torque during the write operation. With these materials breakthroughs, IBM researchers reported a comprehensive evaluation of STT-MRAM devices with PMA materials in 2010. They demonstrated narrow switching voltage distributions and low write-error rate of p-MTJs,

sufficient to fulfill the write requirements of a 64 Mb chip for the first time.[79] This result led to the transition from in-plane MTJs to p-MTJs in STT-MRAM technology. To date, p-MTJs with a CoFeB-based free layer, a MgO tunnel barrier, and a SAF reference layer have been widely adopted in the field and are the building blocks of the most advanced STT-MRAM devices.

STT-MRAM has a unique combination of high speed, high endurance, and non-volatility. Significant investments have been made over the years by major semiconductor companies and tool vendors. The first embedded STT-MRAM product was announced by Samsung in 2019. To date, all major semiconductor foundries including Samsung, TSMC, and Global Foundries have embedded STT-MRAM offering to their customers. Despite the potential that STT-MRAM holds for high speed applications, the first generation of embedded STT-MRAM products was limited to a write speed ranging from 50 to 200 ns and has mainly been used as an eFlash replacement. The ultimate goal for STT-MRAM technology is to expand the application space to replace SRAM as last level cache (LLC). Compared to eFlash replacement, LLC has much more stringent requirements on write speed (< 10 ns), write reliability (write-error-rate (WER) $< 10^{-12}$), and endurance ($> 10^{16}$). For STT-MRAM devices, the switching current increases significantly with shorter (≤ 10 ns) write pulses, due to the conservation of electron spin angular momentum. The STT switching is also a stochastic process and the probability of switching for a given write-pulse width is determined by the write current applied. So more reliable switching also requires larger write current. Experimental data also show that devices written with shorter pulses and to deeper error rate are more prone to WER anomalies, which further

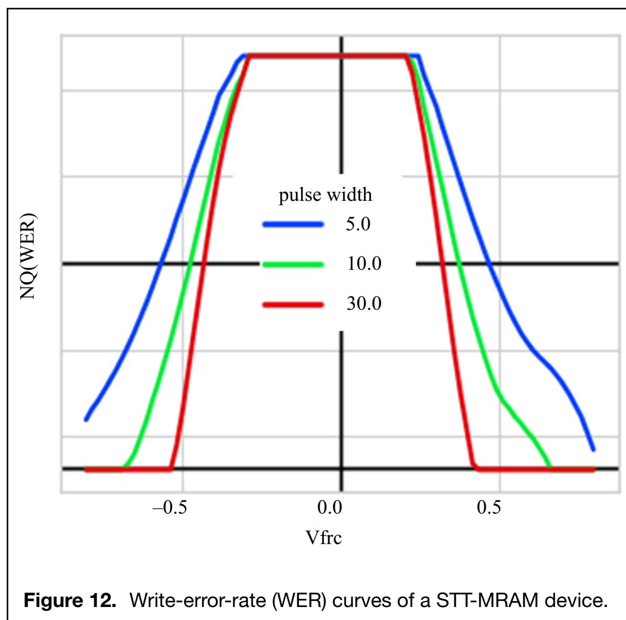


Figure 12. Write-error-rate (WER) curves of a STT-MRAM device.

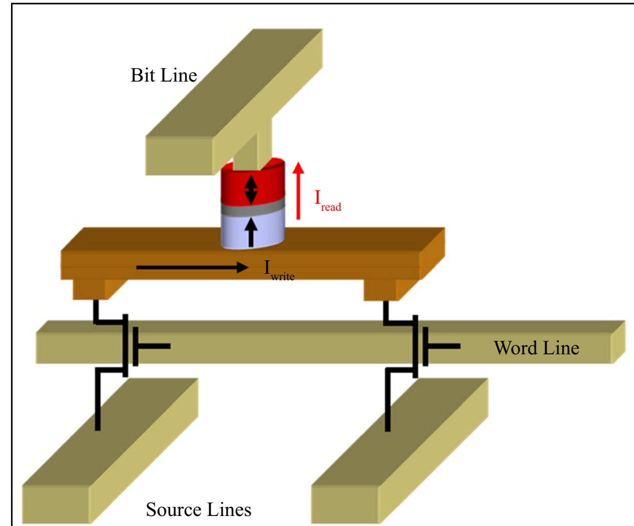


Figure 13. Schematics of a three-terminal SOT-MRAM device; write current flows from source line to word line and read current flows from source line to bit line.

increases the write currents (voltages) (**Figure 12**). During the write process, the dielectric MgO tunnel barrier degrades over time and breaks down eventually. This mechanism sets the endurance limit of STT-MRAM devices—the higher the write current (voltage), the worse the endurance. To date, reliable switching, down to 10^{-11}, with 2–3 ns write pulses was demonstrated in STT-MRAM devices with CoFeB-based free layers.[81] Furthermore, narrow switching voltage distributions that are sufficient to meet the write voltage distribution requirement for LLC applications at 1 × nm nodes were achieved at the same time.[82] Significant reduction of switching current (~ 2X) is still needed, in order to meet the endurance requirement for LLC applications.

Three-terminal spin–orbit torque MRAM (SOT-MRAM) devices have been studied extensively as an alternative to the two-terminal STT-MRAM devices for high speed applications.[83–85] In SOT devices, the write and read paths are separated, as illustrated in **Figure 13**. The dielectric MgO barrier only sees a low read voltage during the operation. Thus, SOT-MRAM is expected to have high endurance. Switching down to a WER = 10^{-5} level was demonstrated with 1 ns write pulses in a single device.[84] People are also actively pursuing a new approach that combines SOT effect with STT effect to improve device performance.[85,86] Compared to two terminal STT-MRAM devices, three-terminal devices significantly reduce the density of MRAM and therefore can eliminate the system-level performance benefit compared to SRAM. It is also not clear whether p-MTJs or in-plane MTJs (i-MTJs) are the right candidates for SOT-MRAM applications with both being pursued in parallel in the field.

To date, there is no convincing demonstration of SOT-MRAM devices outperforming STT-MRAM devices in reliable writing (WER < 10^{-6}) at high speed (< 3 ns). There is

also no array level data to benchmark SOT-MRAM device performance against key requirements for LLC applications, including write speed, write reliability, endurance, retention, etc. These are the gaps that need to be closed in the next few years for SOT-MRAM to be a contender for LLC applications.

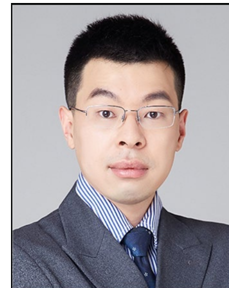
Conclusion

In conclusion, memory technologies with higher density, higher bandwidth, lower power consumption, higher speed, and lower cost are in high demand in the current big data era. SRAM, DRAM, and 3D NAND will still dominate the memory and storage markets in the foreseeable future. Emerging non-volatile memory technologies are promising candidates for eNVM, SCM, near-/in-memory computing applications to overcome the memory wall problem and further improve the performance of computing systems.

References

1. M. Horowitz, *ISSCC Tech. Dig.* (2014), p. 10
2. S. DeBoer, *Symp. VLSI Technol.* (2018), p. 3
3. S. Salahuddin, K. Ni, S. Datta, The era of hyper-scaling in electronics. *Nat. Electron.* **1**, 442 (2018)
4. S. Lai, *Tech. Dig. Int. Electron Devices Meet.* (2003), p. 10
5. R.F. Freitas, W.W. Wilcke, Storage-class memory: The next storage system technology. *IBM J. Res. Dev.* **52**, 439 (2008)
6. G. Servalli, *Tech. Dig. Int. Electron Devices Meet.* (2009), p. 113
7. S. H. Lee, H.C. Park, M.S. Kim, H.W. Kim, M.R. Choi, H.G. Lee, J.W. Seo, S.C. Kim, S.G. Kim, S.B. Hong, S.Y. Lee, J.U. Lee, Y.S. Kim, K.S. Kim, J.I. Kim, M.Y. Lee, H.S. Shin, S.J. Chae, J.H. Song, H.S. Yoon, J.M. Oh, S.K. Min, H.M. Lee, K.R. Hong, J.T. Cheong, S.N. Park, J.C. Ku, Y.S. Sohn, S.K. Park, T.S. Kim, Y.K. Kim, K.W. Park, C.S. Han, W. Kim, H.J. Kim, K.S. Choi, J.H. Lee, S.J. Hong, *Tech. Dig. Int. Electron Devices Meet.* (2011), p. 47
8. M.J. Kang, T.J. Park, Y.W. Kwon, D.H. Ahn, Y.S. Kang, H. Jeong, S.J. Ahn, Y.J. Song, B.C. Kim, S.W. Nam, H.K. Kang, G.T. Jeong, C.H. Chung, *Tech. Dig. Int. Electron Devices Meet.* (2011), p. 39
9. Y. Choi, I. Song, M.-H. Park, H. Chung, S. Chang, B. Cho, J. Kim, Y. Oh, D. Kwon, J. Sunwoo, J. Shin, Y. Rho, C. Lee, M.G. Kang, J. Lee, Y. Kwon, S. Kim, J. Kim, Y.-J. Lee, Q. Wang, S. Cha, S. Ahn, H. Horii, J. Lee, K. Kim, H. Joo, K. Lee, Y.-T. Lee, J. Yoo, G. Jeong, *ISSCC Tech. Dig.* (2012), p. 46
10. D. Kau, S. Tang, I.V Karpov, R. Dodge, B. Klehn, J.A. Kalb, J. Strand, A. Diaz, N. Leung, J. Wu, S.ean Lee, T. Langtry, K. Chang, C. Papagianni, J. Lee, J. Hirst, S. Erra, E. Flores, N. Righos, H. Castro, G. Spadini, *Tech. Dig. Int. Electron Devices Meet.* (2009), p. 617
11. S.R. Ovshinsky, Reversible electrical switching phenomena in disordered structures. *Phys. Rev. Lett.* **21**, 1450 (1968)
12. S.R. Ovshinsky, H. Fritzsche, Amorphous semiconductors for switching, memory, and imaging applications. *IEEE Trans. Electron Devices* **20**, 91 (1973)
13. H.-Y. Cheng, F. Carta, W.-C. Chien, H.-L. Lung, M.J. BrightSky, 3D cross-point phase-change memory for storage-class memory. *J. Phys. D* **52**, 473002 (2019)
14. M.-J. Lee, D. Lee, H. Kim, H.-S. Choi, J.-B. Park, H.G. Kim, Y.-K. Cha, U.-I. Chung, I.-K. Yoo, K. Kim, *Tech. Dig. Int. Electron Devices Meet.* (2012), p. 33
15. D. Garbin, W. Devulder, R. Degraeve, G.L. Donadio, S. Clima, K. Opsomer, A. Fantini, D. Cellier, W.G. Kim, M. Pakala, A. Cockburn, C. Detavernier, R. Delhougne, L. Goux, G.S. Kar, *Tech. Dig. Int. Electron Devices Meet.* (2019), p. 35.1.1
16. H.Y. Cheng, W.C. Chien, I.T. Kuo, E.K. Lai, Y. Zhu, J.L. Jordan-Sweet, A. Ray, F. Carta, F.M. Lee, P.H. Tseng, M. H. Lee, Y.Y. Lin, W. Kim, R. Bruce, C.W. Yeh, C.H. Yang, M. BrightSky, H.L. Lung, *Tech. Dig. Int. Electron Devices Meet.* (2017), p. 28
17. C.W. Yeh, W.C. Chien, R.L. Bruce, H.Y. Cheng, I.T. Kuo, C.H. Yang, A. Ray, H. Miyazoe, W. Kim, F. Carta, E.K. Lai, M. BrightSky, H.L. Lung, *Symp. VLSI Technol.* (2018), p. 205
18. H.Y. Cheng, W.C. Chien, I.T. Kuo, C.W. Yeh, L. Gignac, W. Kim, E.K. Lai, Y.F. Lin, R.L. Bruce, C. Lavoie, C.W. Cheng, A. Ray, F.M. Lee, F. Carta, C.H. Yang, M.H. Lee, H.Y. Ho, M. BrightSky, H.L. Lung, *Tech. Dig. Int. Electron Devices Meet.* (2018), p. 37.3.1
19. H.Y. Cheng, I.T. Kuo, W.C. Chien, C.W. Yeh, Y.C. Chou, N. Gong, L. Gignac, C.H. Yang, C.W. Cheng, C. Lavoie, M. Hopstaken, R.L. Bruce, L. Buzi, E.K. Lai, F. Carta, A. Ray, M.H. Lee, H.Y. Ho, W. Kim, M. BrightSky, H.L. Lung, *Symp. VLSI Technol.* (2020), p. TM1.6
20. T. Kim, H. Choi, M. Kim, J. Yi, D. Kim, S. Cho, H. Lee, C. Hwang, E.-R. Hwang, J. Song, S. Chae, Y. Chun, J.-K. Kim, *Tech. Dig. Int. Electron Devices Meet.* (2018), p. 37.1.1
21. A. Fazio, *Tech. Dig. Int. Electron Devices Meet.* (2020), p. 24.1.1
22. T.W. Hickmott, Low-frequency negative resistance in thin anodic oxide films. *J. Appl. Phys.* **33**, 2669 (1962)
23. H.-S.P. Wong, H.-Y. Lee, S. Yu, Y.-S. Chen, Y. Wu, P.-S. Chen, B. Lee, F.T. Chen, M.-J. Tsai, "Metal–Oxide RRAM." *Proc. IEEE.* **100**, 1951 (2012)
24. M.N. Kozicki, M. Park, M. Mitkova, Nanoscale memory elements based on solid-state electrolytes. *IEEE Trans. Nanotechnol.* **4**, 331 (2005)
25. E. Wu, T. Ando, B. Li, R. Southwick, J. Stathis, Fundamental roles of extreme-value distributions in dielectric breakdown and memory applications (minimum-value versus maximum-value statistics). *Jpn. J. Appl. Phys.* **59**, 0803 (2020)
26. R. Waser, R. Dittmann, G. Staikov, K. Szot, Redox-based resistive switching memories–nanoionic mechanisms, prospects, and challenges. *Adv. Mater.* **21**, 2632 (2009)
27. I.G. Baek, C.J. Park, H. Ju, D.J. Seong, H.S. Ahn, J.H. Kim, M.K. Yang, S.H. Song, E.M. Kim, S.O. Park, C.H. Park, C.W. Song, G.T. Jeong, S. Choi, H.K. Kang, C. Chung, *Tech. Dig. Int. Electron Devices Meet.* (2011), p. 31.8.1
28. H.W. Pan, K.P. Huang, S.Y. Chen, P.C. Peng, Z.S. Yang, C.-H. Kuo, Y.-D. Chih, Y.-C. King, C.J. Lin, *Tech. Dig. Int. Electron Devices Meet.* (2015), p. 10.5.1
29. O. Golonzka, U. Arslan, P. Bai, M. Bohr, O. Baykan, Y. Chang, A. Chaudhari, A. Chen, J. Clarke, C. Connor, N. Das, C. English, T. Ghani, F. Hamzaoglu, P. Hentges, P. Jain, C. Jezewski, I. Karpov, H. Kothari, R. Kotlyar, B. Lin, M. Metz, J. Odonnell, D. Ouellette, J. Park, A. Pirkle, P. Quintero, D. Seghete, M. Sekhar, A. Sen Gupta, M. Seth, N. Strutt, C. Wiegand, H.J. Yoo, K. Fischer, *Symp. VLSI Technol.* (2019), p. T230
30. T. Gokmen, Y. Vlasov, Acceleration of deep neural network training with resistive cross-point devices: Design considerations. *Front. Neurosci.* **10**, 333 (2016)
31. N. Gong, T. Idé, S. Kim, I. Boybat, A. Sebastian, V. Narayanan, T. Ando, Signal and noise extraction from analog memory elements for neuromorphic computing. *Nat. Commun.* **9**, 1 (2018)
32. T. Gokmen, W. Haensch, Algorithm for training neural networks on resistive device arrays. *Front. Neurosci.* **14**, 103 (2020)
33. H. Kim, M. Rasch, T. Gokmen, T. Ando, H. Miyazoe, J.-J. Kim, J. Rozen, S. Kim, Zero-shifting technique for deep neural network training on resistive cross-point arrays. Preprint, arXiv:1907.10228 (2019)
34. I.M. Ross, Semiconductive translating device, US Patent 2,791,760 (1957)
35. H. Ishiwara, Current status and prospects of FET-type ferroelectric memories. *FED J.* **11**, 27 (2000)
36. T.P. Ma, J.-P. Han, Why is nonvolatile ferroelectric memory field-effect transistor still elusive? *IEEE Electron Device Lett.* **23**, 386 (2002)
37. T.S. Böscke, J. Müller, D. Bräuhaus, U. Schröder, U. Böttger, Ferroelectricity in hafnium oxide thin films. *Appl. Phys. Lett.* **99**, 102903 (2011)
38. J. Muller, T.S. Boscke, U. Schroder, S. Mueller, D. Brauhaus, U. Bottger, L. Frey, T. Mikolajick, Ferroelectricity in simple binary ZrO_2 and HfO_2. *Nano Lett.* **12**, 4318 (2012)
39. S. Dunkel, M. Trentzsch, R. Richter, P. Moll, C. Fuchs, O. Gehring, M. Majer, S. Wittek, B. Muller, T. Melde, H. Mulaosmanovic, S. Slesazeck, S. Muller, J. Ocker, M. Noack, D.-A. Lohr, P. Polakowski, J. Muller, T. Mikolajick, J. Hontschel, B. Rice, J. Pellerin, S. Beyer, *Tech. Dig. Int. Electron Devices Meet.* (2017), p. 19.7.1
40. X. Tian, S. Shibayama, T. Nishimura, T. Yajima, S. Migita, A. Toriumi, Evolution of ferroelectric HfO_2 in ultrathin region down to 3 nm. *Appl. Phys. Lett.* **112**, 102902 (2018)
41. X. Lyu, M. Si, X. Sun, M.A. Capano, H. Wang, P.D. Ye, *Symp. VLSI Technol.* (2019), p. T44
42. S.S. Cheema, D. Kwon, N. Shanker, R. dos Reis, S.-L. Hsu, J. Xiao, H. Zhang, R. Wagner, A. Datar, M.R. McCarter, C.R. Serrao, A.K. Yadav, G. Karbasian, C.-H. Hsu, A.J. Tan, L.-C. Wang, V. Thakare, X. Zhang, A. Mehta, E. Karapetrova, R.V. Chopdekar, P. Shafer, E. Arenholz, C. Hu, R. Proksch, R. Ramesh, J. Ciston, S. Salahuddin, Enhanced ferroelectricity in ultrathin films grown directly on silicon. *Nature* **580**, 478 (2020)
43. T.S. Boscke, J. Muller, D. Brauhaus, U. Schroder, U. Bottger, *Tech. Dig. Int. Electron Devices Meet.* (2011), p. 24.5.1
44. N. Gong, T.-P. Ma, *IEEE Electron Device Lett.* **37**, 1123 (2016)
45. W. Chung, M. Si, P.R. Shrestha, J.P. Campbell, K.P. Cheung, P. D. Ye, *Symp. VLSI Technol.* (2018), p. 89
46. M. Si, X. Lyu, P.R. Shrestha, X. Sun, H. Wang, K.P. Cheung, P.D. Ye, Ultrafast measurements of polarization switching dynamics on ferroelectric and anti-ferroelectric hafnium zirconium oxide. *Appl. Phys. Lett.* **115**, 72107 (2019)
47. X. Lyu, M. Si, P.R. Shrestha, K.P. Cheung, P.D. Ye, *Tech. Dig. Int. Electron Devices Meet.* (2019), p. 15.2.1
48. Y.-C. Chiu, C.-H. Cheng, C.-Y. Chang, M.-H. Lee, H.-H. Hsu, S.-S. Yen, *Symp. VLSI Technol.* (2015), p. T184
49. J. Muller, T.S. Boscke, U. Schroder, R. Hoffmann, T. Mikolajick, L. Frey, Nanosecond polarization switching and long retention in a novel MFIS-FET based on ferroelectric HfO_2. *IEEE Electron Device Lett.* **33**, 185 (2012)
50. E. Yurchuk, J. Muller, J. Paul, T. Schlosser, D. Martin, R. Hoffmann, S. Mueller, S. Slesazeck, U. Schroeder, R. Boschke, R. van Bentum, T. Mikolajick, Impact of scaling on the performance of HfO_2-based ferroelectric field effect transistors. *IEEE Trans. Electron Devices* **61**, 3699 (2014)
51. K. Ni, P. Sharma, J. Zhang, M. Jerry, J.A. Smith, K. Tapily, R. Clark, S. Mahapatra, S. Datta, Critical role of interlayer in $Hf_{0.5}Zr_{0.5}O_2$ ferroelectric FET nonvolatile memory performance. *IEEE Trans. Electron Devices* **65**, 2461 (2018)

52. M. Si, X. Lyu, P.D. Ye, Ferroelectric polarization switching of hafnium zirconium oxide in a ferroelectric/dielectric stack. *ACS Appl. Electron. Mater.* **1**, 745 (2019)
53. M. Si, Z. Lin, J. Noh, J. Li, W. Chung, D.Y. Peide, The impact of channel semiconductor on the memory characteristics of ferroelectric field-effect transistors. *IEEE J. Electron Devices Soc.* **8**, 846 (2020)
54. M. Si, P.D. Ye, The critical role of charge balance on the memory characteristics of ferroelectric field-effect transistors. *IEEE Trans. Electron Devices* **68**, 5108 (2021)
55. K. Toprasertpong, M. Takenaka, S. Takagi, *Tech. Dig. Int. Electron Devices Meet.* (2019), p. 23.7.1
56. M. Si, J. Andler, X. Lyu, C. Niu, S. Datta, R. Agrawal, P.D. Ye, Indium–tin-oxide transistors with one nanometer thick channel and ferroelectric gating. *ACS Nano* **14**, 11542 (2020)
57. A.A. Sharma, B. Doyle, H.J. Yoo, I.-C. Tung, J. Kavalieros, M.V Metz, M. Reshotko, P. Majhi, T. Brown-Heft, Y.-J. Chen, V.H. Le, *Tech. Dig. Int. Electron Devices Meet.* (2020), p. 18.5.1
58. S. Dutta, H. Ye, W. Chakraborty, Y.-C. Luo, M.S. Jose, B. Grisafe, A. Khanna, I. Lightcap, S. Shinde, S. Yu, S. Datta, *Tech. Dig. Int. Electron Devices Meet.* (2020), p. 36.4.1
59. K. Florent, S. Lavizzari, L. Di Piazza, M. Popovici, E. Vecchio, G. Potoms, G. Groeseneken, J. Van lHoudt, *Symp. VLSI Technol.* (2017), p. T158
60. M.-K. Kim, I.-J. Kim, J.-S. Lee, CMOS-compatible ferroelectric NAND flash memory for high-density, low-power, and high-speed three-dimensional memory. *Sci. Adv.* **7**, eabe1341 (2021)
61. M. Jerry, P.-Y. Chen, J. Zhang, P. Sharma, K. Ni, S. Yu, S. Datta, *Tech. Dig. Int. Electron Devices Meet.* (2017), p. 6.2.1
62. W. Chung, M. Si, P.D. Ye, *Tech. Dig. Int. Electron Devices Meet.* (2018), p. 15.2.1
63. Z. Wang, B. Crafton, J. Gomez, R. Xu, A. Luo, Z. Krivokapic, L. Martin, S. Datta, A. Raychowdhury, A.I. Khan, *Tech. Dig. Int. Electron Devices Meet.* (2018), p. 13.3.1
64. A. Keshavarzi, K. Ni, W. Van Den Hoek, S. Datta, A. Raychowdhury, Ferroelectronics for edge intelligence. *IEEE Micro* **40**, 33 (2020)
65. J.C. Slonczewski, Conductance and exchange coupling of two ferromagnets separated by a tunneling barrier. *Phys. Rev. B* **39**, 6995 (1989)
66. J.C. Slonczewski, Current-driven excitation of magnetic multilayers. *J. Magn. Magn. Mater.* **159**, L1 (1996)
67. L. Berger, Emission of spin waves by a magnetic multilayer traversed by a current. *Phys. Rev. B* **54**, 9353 (1996)
68. J.Z. Sun, Current-driven magnetic switching in manganite trilayer junctions. *J. Magn. Magn. Mater.* **202**, 157 (1999)
69. J.A. Katine, F.J. Albert, R.A. Buhrman, E.B. Myers, D.C. Ralph, Current-driven magnetization reversal and spin-wave excitations in Co/Cu/Co pillars. *Phys. Rev. Lett.* **84**, 3149 (2000)
70. T. Miyazaki, N. Tezuka, Giant magnetic tunneling effect in Fe/Al_2O_3/Fe junction. *J. Magn. Magn. Mater.* **139**, L231 (1995)
71. J.S. Moodera, L.R. Kinder, T.M. Wong, R. Meservey, Large magnetoresistance at room temperature in ferromagnetic thin film tunnel junctions. *Phys. Rev. Lett.* **74**, 3273 (1995)
72. W.H. Butler, X.-G. Zhang, T.C. Schulthess, J.M. MacLaren, Spin-dependent tunneling conductance of Fe|MgO|Fe sandwiches. *Phys. Rev. B* **63**, 54416 (2001)
73. S. Yuasa, T. Nagahama, A. Fukushima, Y. Suzuki, K. Ando, Giant room-temperature magnetoresistance in single-crystal Fe/MgO/Fe magnetic tunnel junctions. *Nat. Mater.* **3**, 868 (2004)
74. S.S.P. Parkin, C. Kaiser, A. Panchula, P.M. Rice, B. Hughes, M. Samant, S.-H. Yang, Giant tunnelling magnetoresistance at room temperature with MgO (100) tunnel barriers. *Nat. Mater.* **3**, 862 (2004)
75. S. Ikeda, J. Hayakawa, Y. Ashizawa, Y.M. Lee, K. Miura, H. Hasegawa, M. Tsunoda, F. Matsukura, H. Ohno, Tunnel magnetoresistance of 604% at 300 K by suppression of Ta diffusion in CoFeB/MgO/CoFeB pseudo-spin-valves annealed at high temperature. *Appl. Phys. Lett.* **93**, 82508 (2008)
76. J.Z. Sun, Spin-current interaction with a monodomain magnetic body: A model study. *Phys. Rev. B.* **62**, 570 (2000)
77. D.C. Worledge, G. Hu, P.L. Trouilloud, D.W. Abraham, S. Brown, M.C. Gaidis, J. Nowak, E.J. O'Sullivan, R.P. Robertazzi, J.Z. Sun, W.J. Gallagher, *Tech. Dig. Int. Electron Devices Meet.* (2010), p. 12.5.1
78. S. Ikeda, K. Miura, H. Yamamoto, K. Mizunuma, H.D. Gan, M. Endo, S. Kanai, J. Hayakawa, F. Matsukura, H. Ohno, A perpendicular-anisotropy CoFeB-MgO magnetic tunnel junction. *Nat. Mater.* **9**, 721 (2010)
79. D.C. Worledge, G. Hu, D.W. Abraham, J.Z. Sun, P.L. Trouilloud, J. Nowak, S. Brown, M.C. Gaidis, E.J. O'sullivan, R.P. Robertazzi, Spin torque switching of perpendicular Ta|CoFeB|MgO-based magnetic tunnel junctions. *Appl. Phys. Lett.* **98**, 22501 (2011)
80. G. Hu, T. Topuria, P.M. Rice, J. Jordan-Sweet, D.C. Worledge, Optimization of tunneling magnetoresistance in perpendicular magnetic tunnel junctions with Co|Pd reference layers. *IEEE Magn. Lett.* **4**, 3000104 (2013)
81. G. Hu, D. Kim, J. Kim, C. Kothandaraman, G. Lauer, H.K. Lee, N. Marchack, M. Reuter, R.P. Robertazzi, J.Z. Sun, T. Suwannasiri, J.J. Nowak, P.L. Trouilloud, S. Woo, D.C. Worledge, M.G. Gottwald, S.L. Brown, B. Doris, C.P. D'Emic, P. Hashemi, D. Houssameddine, Q. He, *Tech. Dig. Int. Electron Devices Meet.* (2019), p. 2.6.1
82. E.R.J. Edwards, G. Hu, S.L. Brown, C.P. D'Emic, M.G. Gottwald, P. Hashemi, H. Jung, J. Kim, G. Lauer, J.J. Nowak, J.Z. Sun, T. Suwannasiri, P.L. Trouilloud, S. Woo, D.C. Worledge, *Tech. Dig. Int. Electron Devices Meet.* (2020), p. 24.4.1
83. H. Honjo, T.V.A. Nguyen, T. Watanabe, T. Nasuno, C. Zhang, T. Tanigawa, S. Miura, H. Inoue, M. Niwa, T. Yoshiduka, Y. Noguchi, M. Yasuhira, A. Tamakoshi, M. Natsui, Y. Ma, H. Koike, Y. Takahashi, K. Furuya, H. Shen, S. Fukami, H. Sato, S. Ikeda, T. Hanyu, H. Ohno, T. Endoh, *Tech. Dig. Int. Electron Devices Meet.* (2019), p. 28.5.1
84. K. Garello, F. Yasin, H. Hody, S. Couet, L. Souriau, S.H. Sharifi, J. Swerts, R. Carpenter, S. Rao, W. Kim, J. Wu, K.K.V. Sethu, M. Pak, N. Jossart, D. Crotti, A. Furnemont, G.S. Kar, *Symp. VLSI Technol.* (2019), p. T194
85. M. Wang, W. Cai, D. Zhu, Z. Wang, J. Kan, Z. Zhao, K. Cao, Z. Wang, Y. Zhang, T. Zhang, C. Park, J.-P. Wang, A. Fert, W. Zhao, Field-free switching of a perpendicular magnetic tunnel junction through the interplay of spin–orbit and spin-transfer torques. *Nat. Electron.* **1**, 582 (2018)
86. N. Sato, G.A. Allen, W.P. Benson, B. Buford, A. Chakraborty, M. Christenson, T.A. Gosavi, P.E. Heil, N.A. Kabir, B.J. Krist, K.P. O'Brien, K. Oguz, R.R. Patil, J. Pellegren, A.K. Smith, E.S. Walker, P. J. Hentges, M.V. Metz, M. Seth, B. Turkot, C.J. Wiegand, H.J. Yoo, I.A. Young, *Symp. VLSI Technol.* (2020), p. 1 □



Mengwei Si received his BS degree in electronic engineering from Shanghai Jiao Tong University, Shanghai, China, in 2012, and a PhD degree in electrical and computer engineering from Purdue University in 2018. He was a postdoc with the Department of Electrical and Computer Engineering at Purdue University between 2018 and 2021. Si is currently an associate professor with the Department of Electronic Engineering, Shanghai Jiao Tong University. His research interests include semiconductor fabrication, semiconductor materials and devices, ferroelectric materials, nanoscale devices and technology, and atomic layer deposition. Si can be reached by email at mengwei.si@sjtu.edu.cn.

Huai-Yu Cheng received her BS, MS, and PhD degrees in materials science and engineering from National Tsing Hua University, Taiwan, in 2001, 2003, and 2007, respectively. She joined Macronix International Co., Ltd. in 2007, and the IBM/Macronix PCRAM Joint Project in 2008, currently based in Yorktown Heights, New York. Cheng has engaged in research and advanced technology development for emerging memories including phase change memory and selector devices for three-dimensional cross-point and storage-class memory. She has published more than 70 conference and journal papers including 12 invited papers, two review articles, one book chapter, and has been granted over 20 US Patents. Cheng can be reached by email at hymcheng@us.ibm.com or hycheng@mxic. com.tw.

Takashi Ando received his BS and ME degrees from the University of Tokyo in 1999 and 2001, and his PhD degree from Osaka University in 2010. After joining IBM Research, he worked on early research to productization phases of High-k Metal Gate Technology. He is currently leading the global team to develop new analog memories for AI applications. He is a recipient of the Japan Society of Applied Physics Young Scientist Award in 2011 and IEEE EDS George E. Smith Award in 2013. Ando can be reached by email at andot@us.ibm.com.

Guohan Hu is a distinguished research staff member and manager of the magnetic and randomaccess memory (MRAM) Materials and Devices Group at the IBM T.J. Watson Research Center. She received her PhD degree in materials science and engineering from Cornell University in 2002. Guohan's research has focused on magnetic oxide thin films, patterned magnetic recording media, and MRAM materials and devices. Hu can be reached by email at hug@us.ibm.com.

Peide D. Ye is the Richard J. and Mary Jo Schwartz Chair Professor of Electrical and Computer Engineering at Purdue University. He received his PhD degree from the Max Planck Institute of Solid State Research, Stuttgart, Germany, in 1996. Before joining Purdue faculty in 2005, Ye worked for NTT, NHMFL/Princeton University, and Bell Labs/Lucent Technologies/Agere Systems. His current research focuses on atomic layer deposition high-k integration on novel channel materials including III–V, Ge, complex oxides, and two-dimensional crystals. He authored and co-authored more than 300 peer-reviewed articles and 500 conference presentations. Ye is a Fellow of IEEE and the American Physical Society. He is also recognized as a Highly Cited Researcher among 6000 worldwide in all fields.Ye can be reached by email at yep@purdue.edu.