Solid-State Electronics 125 (2016) 25–38

Contents lists available at ScienceDirect

Solid-State Electronics

journal homepage: www.elsevier.com/locate/sse

A review of emerging non-volatile memory (NVM) technologies and applications

An Chen

IBM Research, San Jose, CA 95120, USA

ARTICLE INFO

Article history:
Available online 30 July 2016

The review of this paper was arranged by Jurriaan Schmitz

Keywords:
Nonvolatile memory
PCM
STTRAM
RRAM
FeFET
Storage
Memory hierarchy
Emerging architecture
Selector
Neuromorphic computing
Hardware security

ABSTRACT

This paper will review emerging non-volatile memory (NVM) technologies, with the focus on phase change memory (PCM), spin-transfer-torque random-access-memory (STTRAM), resistive random-access-memory (RRAM), and ferroelectric field-effect-transistor (FeFET) memory. These promising NVM devices are evaluated in terms of their advantages, challenges, and applications. Their performance is compared based on reported parameters of major industrial test chips. Memory selector devices and cell structures are discussed. Changing market trends toward low power (*e.g.*, mobile, IoT) and data-centric applications create opportunities for emerging NVMs. High-performance and low-cost emerging NVMs may simplify memory hierarchy, introduce non-volatility in logic gates and circuits, reduce system power, and enable novel architectures. Storage-class memory (SCM) based on high-density NVMs could fill the performance and density gap between memory and storage. Some unique characteristics of emerging NVMs can be utilized for novel applications beyond the memory space, *e.g.*, neuromorphic computing, hardware security, *etc*. In the beyond-CMOS era, emerging NVMs have the potential to fulfill more important functions and enable more efficient, intelligent, and secure computing systems.

1. Introduction

Temporary and permanent data storage is required in any functional information processing systems, which has so far been fulfilled by CMOS-based memories, *i.e.*, SRAM, DRAM, and Flash memory. The speed gap between logic and memory has become a critical system performance bottleneck, *i.e.*, the "memory wall" [1]. Hierarchical memory systems made from devices with varying speed, density and cost have been adopted to optimize the performance-cost tradeoff. With CMOS scaling approaching fundamental limits, some novel non-volatile memory (NVM) concepts have been proposed and made significant progress in recent years [2]. Although high-performance computing is still an important driver for semiconductor technology innovation, consumer electronics is shifting toward mobile, pervasive connectivity, and data-centric applications. The changing market trend imposes different requirements on hardware, *e.g.*, ultra-low power computing, high-density and low-cost data storage, novel functions, *etc*. Emerging NVMs with high performance, good scalability, and new functionalities may become important technology enablers to fulfill these requirements.

2. Emerging NVM device options

2.1. Memory taxonomy and emerging NVM candidates

Fig. 1 shows the memory taxonomy from the 2013 International Technology Roadmap for Semiconductors (ITRS) Emerging Research Devices (ERD) chapter [2]. NVMs with prototype test chips or early production are included in the "prototypical" category, which covers ferroelectric random-access-memory (FeRAM), phase change memory (PCM), magnetic RAM (MRAM), and spin-transfer-torque RAM (STTRAM). However, some of these prototypical memories may still be considered "emerging" in the research community, with significant R&D activities. Therefore, this paper will start with a broad scope of emerging memories including devices in both "prototypical" and "emerging" categories in Fig. 1.

Emerging NVMs often involve novel mechanisms and materials different from those of mature memories based on Si CMOS. These materials include ferroelectric dielectrics, ferromagnetic metals, chalcogenides, transitional metal oxides, carbon materials, *etc*.

E-mail address: an.chen@aya.yale.edu

http://dx.doi.org/10.1016/j.sse.2016.07.006
0038-1101/

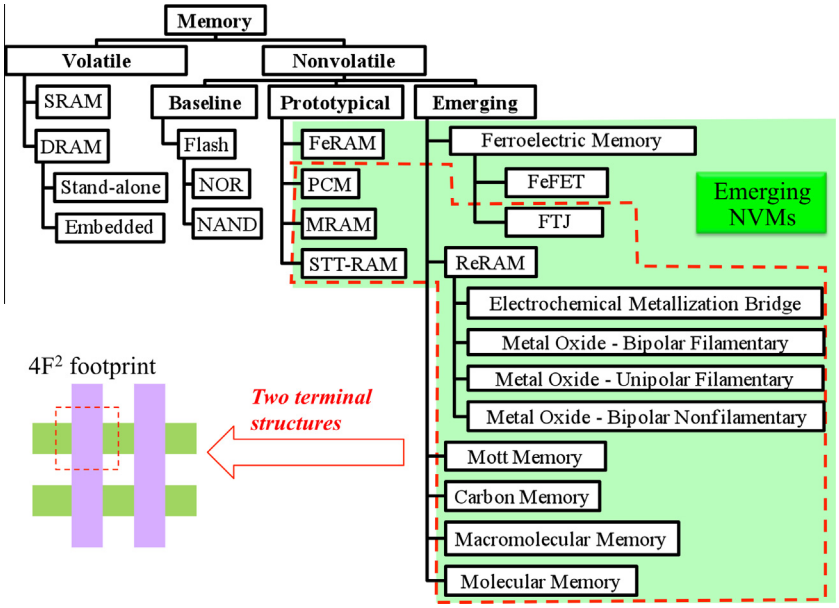

Fig. 1. Memory taxonomy from the 2013 ITRS Emerging Research Devices (ERD) chapter [2]. Many emerging NVMs have a simple two-terminal structure, suitable for high-density crossbar memory arrays.

Their switching mechanisms extend beyond classical electronic processes, to quantum mechanical phenomena, ionic reactions, phase transition, molecular reconfiguration, *etc.* Most of the emerging NVMs are based on two-terminal switching elements, which are suitable for high-density memory architectures, *e.g.*, cross-bar arrays (CBAs).

With the large variety of emerging NVMs, it is necessary to compare their performance and evaluate their long-term potential. ERD conducts a regular survey-based critical review of devices in the "emerging" category, using eight criteria and a 1–3 score (with 3 representing the best and 1 the worst). The result of the 2013 ERD critical review is summarized in Fig. 2. Good scalability is considered one of the key advantages of emerging NVMs, while reliability is a common challenge. RRAM stands out as the most promising device among the six candidates in Fig. 2, followed by FeFET memory. "Carbon-based memory" involves mixed types of carbon materials (*e.g.*, carbon nanotube, graphene, amorphous carbon, *etc.*) and has attracted growing interest [3–5]. "Molecular memory" is pursued mostly for ultimate scalability, while its performance is relatively poor [6,7]. Progress has been made on "Mott memory" based on metal–insulator transition, but materials with high transition temperature are needed for practical applications [8–10]. "Macromolecular memory" may be useful in flexible electronics and low-cost applications, but has limitations in performance and reliability [11,12].

In a recent workshop organized by ERD in 2014, nine emerging NVM technologies were evaluated based on both demonstrated performance ("most promising") and foreseen potential ("most need of resources"), as shown in Fig. 3. PCM, STTRAM, and RRAM (including both oxide based RRAM and conductive-bridge RAM) are among the top candidates with the most promising performance. Emerging ferroelectric memory, especially FeFET with doped ferroelectric HfO_x, is considered the top choice that needs more resources to explore its long-term potential. The rest of the paper will focus on these four NVM devices: PCM, STTRAM, RRAM, and FeFET memory.

PCM, STTRAM, and RRAM are all back-end-of-line (BEOL) memories with two-terminal memory elements integrated between two metal layers. The structures of these memory elements are shown in Fig. 4. Behind the simplicity on the appearance, each memory involves complex issues in stack engineering, thin-film deposition, process optimization, and device designs, which will be explained in details in the following sections. PCM switching is based on Joule heating induced phase transition in chalco-

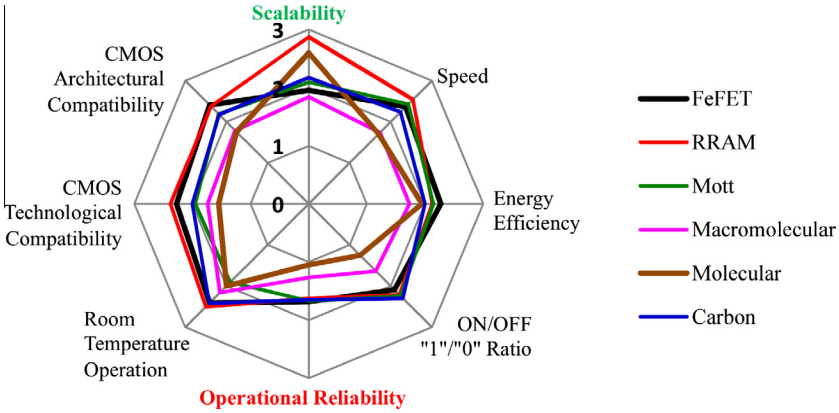

Fig. 2. 2013 ITRS ERD critical review of emerging memories based on eight criteria, where "3" and "1" represent the best and the worst assessment scores, respectively.

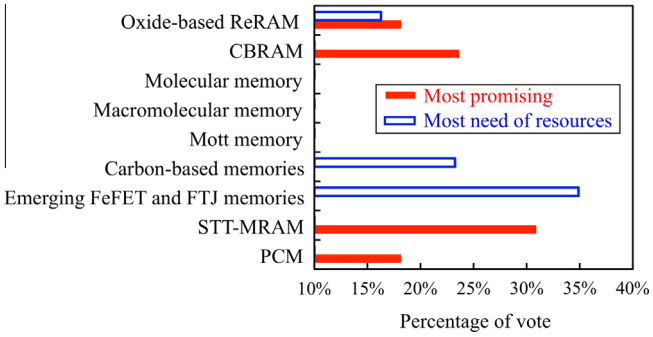

Fig. 3. Survey of the "most promising" emerging NVM device candidates (red solid bars) in the 2014 ERD Emerging Memory Assessment Workshop (Albuquerque, NM). The "most need of resources" (blue empty bars) refers to less mature devices with interesting properties, which need more resources to prove their feasibility. Only vote above 10% is shown in the figure. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

genides, which is unipolar (*i.e.*, switching on and off in the same bias direction). STTRAM switching involves direction-dependent spin-transfer-torque of spin-polarized electrons and is therefore bipolar (*i.e.*, switching on and off in the opposite bias directions). Both unipolar and bipolar switching phenomena have been observed in RRAM, although the latter has generally demonstrated more stable switching behaviors and better performance.

2.2. Phase change memory (PCM)

PCM is based on reversible transition between crystalline phase (low resistance) and amorphous phase (high resistance) of chalcogenides [13]. The transition from amorphous to crystalline phase (*i.e.*, a set process) determines PCM performance (speed) while the reverse process (*i.e.*, a reset process) is the power-limiting step. PCM performance depends critically on the property of phase change materials and memory cell design. The microscopic mechanisms of phase change in Ge-Sb-Te (GST) alloys have been thoroughly studied, which help to guide material engineering to improve device characteristics [14]. PCM cells need to be designed to simplify processing, optimize power efficiency (*i.e.*, improve thermal isolation), and reduce reset current. Sub-lithographic cell design has been developed, through contact-minimized or volume-minimized approaches, to lower switching power. Overall, phase change materials demonstrate desirable scaling behaviors, *e.g.*, phase transition at highly scaled dimension (several nm in thin-film thickness or nano-particle diameter), higher crystallization temperature (*i.e.*, longer retention) at smaller dimension, decreasing thermal conductivity (*i.e.*, higher power efficiency) with thinner films, linear dependence of threshold voltage on device size, improved endurance at smaller dimension, *etc.* PCM cell size is limited mainly by access devices (or selector devices) due to the relatively large current required for switching. Bipolar junction transistor (BJT), vertical transistor, and even diode have been experimented as access devices to reduce PCM cell size. PCM failures could be caused by void formation at electrodes, change in mass density, elemental segregation, *etc.*

PCM is among the most developed novel NVMs and has demonstrated promising performance (*e.g.*, <100 ns switching speed, $>10^9$ cycle endurance) [15]. However, its initial target of replacing incumbent memories (*e.g.*, Flash or even DRAM) has not been achieved due to both continuous improvement of incumbent memories and limitations of PCM itself (*e.g.*, cost) [16]. Although it is relatively easy for PCM to meet standalone Flash memory requirements, achieving both high speed (~ns) and long endurance ($>10^{12}$ cycles) for working memories is still very challenging for PCM. Atomic-level engineering by doping GST materials may improve both speed and endurance [17]. Applications with high reliability targets (*e.g.*, long retention at high temperature for automotive) would require much higher crystallization temperature. PCM has also demonstrated the potential for novel applications, *e.g.*, storage class memory, ternary content addressable memory (TCAM), neuromorphic computing, *etc.*

2.3. Spin-transfer-torque random-access-memory (STTRAM)

STTRAM improves the writing mechanism of conventional field-switching MRAM with spin transfer torques, which is more efficient and more scalable. The memory element is the same magnetic tunnel junction (MTJ) with two ferromagnetic layers separated by an ultra-thin oxide barrier (typically 1–2 nm thick MgO). The parallel and anti-parallel orientations of the two ferromagnetic layers contribute to low and high resistance states, respectively. STTRAM with in-plane magnetic anisotropy has entered early commercialization (64 Mb with 90 nm CMOS process) [18], while further scaling has focused on MTJs with perpendicular magnetic anisotropy (PMA) [19]. Among all NVMs, STTRAM has demonstrated the highest performance (measured by <10 ns write speed and $>10^{12}$ cycle endurance) and is the most promising candidate for embedded NVM applications. Although STTRAM bit-cell size is much larger than MTJ size (due to the large size of access transistors required by high switching current), STTRAM is still smaller than SRAM. With nonvolatility, STTRAM can also save standby power and simplify the memory hierarchy.

STTRAM still faces some critical challenges. With small on/off ratio (measured by tunneling magneto-resistance ratio, or TMR ratio), STTRAM needs well-designed reading schemes and is sensitive to increasing MTJ variability with scaling. Reducing MTJ size and switching current while maintaining sufficient thermal

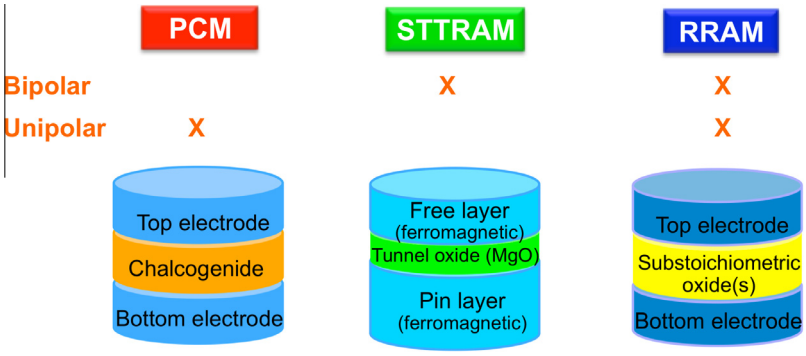

Fig. 4. Structure and operation polarity of PCM, STTRAM, and RRAM.

stability (*i.e.*, retention) requires device and material innovations, *e.g.*, PMA, dual-MgO structure, composite free-layer, *etc.* The distributions of read voltage, write voltage and breakdown voltage have to be separated with sufficient margins in large arrays [20]. Embedded memories have to meet stringent BEOL thermal budget (typically 400 °C for $\geqslant$60 min), which is still an active target for STTRAM R&D [21]. STTRAM is highly sensitive to the fabrication process, *e.g.*, substrate smoothness, etching damage, encapsulation, *etc.* Choice of seed layer and capping layer affects MTJ properties. Optimization of the magnetic and electrical properties of MTJs with many layers of ferromagnetic thin-films, non-magnetic metals, and ultra-thin oxide relies on well-understood device physics and carefully controlled material engineering.

Some recently reported spintronic phenomena may further improve STTRAM design and performance. A discovery of large change of magnetic anisotropy in a Fe(001)/MgO(001) junction by a small electric field may help to reduce STTRAM switching power, which is known as voltage controlled magnetic anisotropy (VCMA) [22]. A giant spin Hall effect (SHE) discovered in heavy metals (*e.g.*, Ta, W) may provide a more efficient source of spin torque to switch MTJs and enable separate writing and reading paths to improve STTRAM reliability [23]. These discoveries offer new opportunities to continue the advancement of STTRAM technologies.

2.4. Resistive random-access-memory (RRAM)

RRAM typically refers to the electrical switching between different resistance states observed in numerous metal oxides (*e.g.*, NiO_x, HfO_x, TiO_x, TaO_x, $Pr_xCa_{1-x}MnO_3$), although similar phenomenon has also been reported in non-oxide materials (*e.g.*, silicon, sulfides, chalcogenides). An early report of RRAM based on simple binary metal oxide attracted great attention, owing to promising scalability, CMOS compatibility, and sufficient performance for post-Flash NVM solutions [24]. Interest in RRAM was further enhanced by the discovery of resistive switching in ALD HfO_x, usually with reactive caps to create sub-stoichiometric oxides with oxygen vacancies needed for switching [25]. Recent focus is shifting to materials with more stable switching behaviors (*e.g.*, TaO_x [26]) and multi-oxide stack with multiple optimization knobs (*e.g.*, oxide compositions, interfaces [27]). A 32 Gb RRAM test chip was demonstrated using 24 nm CMOS technology [28].

Conductive-Bridge RAM (CBRAM) is a type of RRAM with a reactive electrode that supplies mobile ions (*e.g.*, Cu^+, Ag^+) to migrate across insulating dielectrics (*a.k.a.* solid-state electrolytes) and form metallic filaments in the on-state [29]. Various types of solid-state electrolytes have been tested for CBRAM, including oxides, sulfides, chalcogenides, α-Si, *etc.* CBRAM has demonstrated large on/off ratio, fast speed, and long endurance, although retention still needs improvement. A 16 Gb CBRAM test chip has been reported recently [30].

Tradeoffs exist among key RRAM parameters, *e.g.*, speed-retention, power-speed, endurance-retention, *etc.* A major challenge of RRAM is reliability (*e.g.*, endurance, retention), variability, and failure mechanisms. RRAM can typically achieve endurance over 10^6 cycles [24]. Although longer endurance has been reported (e.g., 10^9 cycles) [31], reliable endurance testing is constrained by selected verification. Some RRAM devices have shown rapid "relaxation" behavior (*i.e.*, fast retention loss) immediately after programming, which is detrimental to memory stability and retention [32]. RRAM switching is generally attributed to the formation and rupture of conductive filaments inside of insulating oxides. The localized switching and transport mechanisms contribute to the scaling advantage of RRAM. On the other hand, the location, dimension, and composition of the filaments vary from cycle to cycle and from cell to cell, contributing to an intrinsically stochastic switching process. The stochastic RRAM switching mechanisms are reflected in the large variation in device resistance and switching voltages [33]. The failure mechanisms of RRAM are still not well understood and random failures may occur unpredictably. Due to these random behaviors, statistical characterization and analysis tools have been increasingly utilized in RRAM research. RRAM switching is similar to a controllable and reversible soft dielectric breakdown process. Therefore, dielectric reliability and breakdown models have been frequently applied on RRAM simulation [34]. Many RRAM devices require a forming process (under higher bias or for longer duration than set/reset conditions) before stable switching can be achieved. Forming plays a critical role in defining RRAM switching characteristics [35]; however, this one-time process may complicate RRAM design and operation. It is not yet clear whether truly "forming-free" RRAM can be achieved.

2.5. Ferroelectric-FET (FeFET) memory

In FeFET, a ferroelectric layer is used as the gate dielectrics of a FET, where the change of ferroelectric polarization modulates FET channel conductance. Since FeFET switching is field-driven at transistor gate where the leakage current is minimal, FeFET is expected to have very low switching power. FeFET is similar to Flash memory but with the floating gate for data storage replaced by ferroelectric polarization. FeFET is not a new concept, but its development has been hindered by depolarization field, gate leakage, and the lack of scalable ferroelectric gate dielectrics [36]. A recent discovery of ferroelectricity in doped HfO_x has reignited interest in FeFET [37]. Integration of Si-doped HfO_x in 28 nm HKMG process has been demonstrated. Switching speed is as fast as 20 ns, although switching voltage is still relatively high (*e.g.*, 5 V). The endurance of ferroelectric HfO_x in FeFET is currently limited to 10^4–10^5 cycles possibly due to parasitic charge trapping effects, but longer endurance has been achieved in capacitor structures (10^8–10^9 cycles). Unlike BEOL memories, FeFET is fabricated in the front-end-of-line (FEOL) with more stringent material requirements. Ferroelectric HfO_x may significantly improve the scalability of FeFET. With full CMOS-compatibility, simple 1-transistor (1T) structure, and the performance close to that of DRAM, HfO_x-based FeFET may provide another promising embedded NVM candidate.

The "ferroelectric memory" in Fig. 1 includes both FeFET and FTJ (ferroelectric tunnel junction). FTJ is a two-terminal device with a thin ferroelectric layer sandwiched between two metal electrodes [38]. The tunneling resistance can be modulated by ferroelectric polarization reversal, introducing different resistance states for memory application. FTJ is less mature than FeFET and not discussed in details in this paper.

2.6. Comparison of major emerging NVMs

Table 1 summarizes the main advantages and challenges of the four major emerging NVMs discussed above. FeFET is the only emerging NVM with 1-transistor structure among them. It offers a promising low-power NVM solution. PCM is the most mature emerging NVM with proven performance and would benefit from further reduction of switching power. STTRAM achieves the highest performance and significant R&D efforts have focused on material and process optimization for its early commercialization. RRAM has advantages in simplicity and potentially low-cost, but reliability needs further improvement.

The performance of memory devices can be evaluated based on scalability, speed, power, and reliability, as shown in Fig. 5. Each category involves multiple factors. Scalability is not only determined by how small a functional memory element can be made but also affected by memory cell structures and feasibility of 3D

Table 1
Summary of advantages and challenges of major emerging NVMs.

	Main advantages	Key challenges
FeFET	• 1T cell structure • Low-power field-driven • High performance • Ferroelectric doped HfO_x	• Material and processing • FEOL integration • Reliability and parasitic effects (*e.g.*, charge trapping)
PCM	• Maturity • Proven performance	• Reliability • Disturbance • High switching power
STTRAM	• High performance • Well-understood physics • Novel mechanisms (*e.g.*, SHE, VCMA) to extend capabilities	• Reducing I_c/Δ (power-stability tradeoff) • MTJ patterning and etching • BEOL thermal budget
RRAM	• Simplicity and low cost • High density • Versatile materials, structures, and behaviors	• Reliability and failures • Stochastic mechanism and intrinsic variability • Forming

architectures. Multi-level cell (MLC) capability increases the data density with the same physical cell size. Writing speed depends on the switching mechanisms, while reading speed is determined by the sensing scheme and the memory on/off ratio. Memory array layout and parasitics also have significant impact on memory speed. Power is determined by operation voltage and current. It should be mentioned that low power does not necessarily mean low energy, *e.g.*, Flash memories have very small write current (due to tunneling mechanism) and therefore low power, but their long write time increases the write energy. Reliability (*e.g.*, retention, endurance, *etc.*) and variability impact the feasibility of large memory array and products. In Fig. 5, PCM, STTRAM, and RRAM are compared qualitatively with green, yellow, and red icons indicating "best", "medium", and "worst", respectively. Such comparison should be considered subjective and dependent on applications, and therefore is for reference only.

3. Memory cell structure and selector devices

A functional NVM cell consists of a storage element (*i.e.*, a non-volatile on/off switch) and a selector (to select the element in a memory array). Fig. 6 illustrates different memory cell structures and their suitability for planar and 3D memory architectures.

Flash memory is a floating gate (as the storage element) and a transistor (as the selector) combined in one device. FeFET memory has a 1-transistor (1T) structure like Flash memory. It is yet unknown whether it is possible to build vertical FeFET memory like 3D Flash. Two-terminal NVM devices can work with either transistor selectors in a 1-transistor-1-resistor (1T1R) cell or two-terminal selectors in a 1-selector-1-resistor (1S1R) cell. RRAM with sufficient self-selecting properties (*e.g.*, intrinsic rectifying or nonlinear behaviors) may enable functional memories without external selectors in a 1-resistor (1R) cell [39]. Integrating external selectors on the sidewall in 3D vertical memory architectures is very challenging; therefore, 1T1R and 1S1R cells of PCM and STTRAM are most likely only feasible for planar and 3D stackable architectures. Self-selecting 1R cells are more suitable for 3D vertical architectures, but optimizing both memory and selector properties in one device is challenging. Transistors requiring high processing temperature are unsuitable as selectors for 3D stackable memories. Therefore, low-temperature two-terminal selectors are key enablers for 3D stackable memories. In principle, STTRAM can be built in 3D stackable structures, but in reality, the stringent smoothness requirement for high-performance MTJs may limit STTRAM stacking. Although 3D stackable architectures can achieve high bit density on the same footprint, it is not bit-cost scalable like 3D vertical architectures because every stacked layer requires lithography and patterning steps.

Transistors provide not only selector functions but also switching control (*e.g.*, adjustable current compliance), which is important for many RRAM devices. Two-terminal selectors usually do not provide such control other than fixed series resistance. To avoid excessive damage, RRAM cells with self-compliance (typically through internal series resistance) are needed in memory arrays without external compliance.

Two-terminal selectors for high-density CBAs could be rectifying diodes, nonlinear devices, or volatile switches, as shown in Fig. 7 [40]. These devices utilize either asymmetry or nonlinearity in their characteristics to suppress sneak leakage in CBAs [41]. The unselected junctions in CBAs usually have lower bias than the

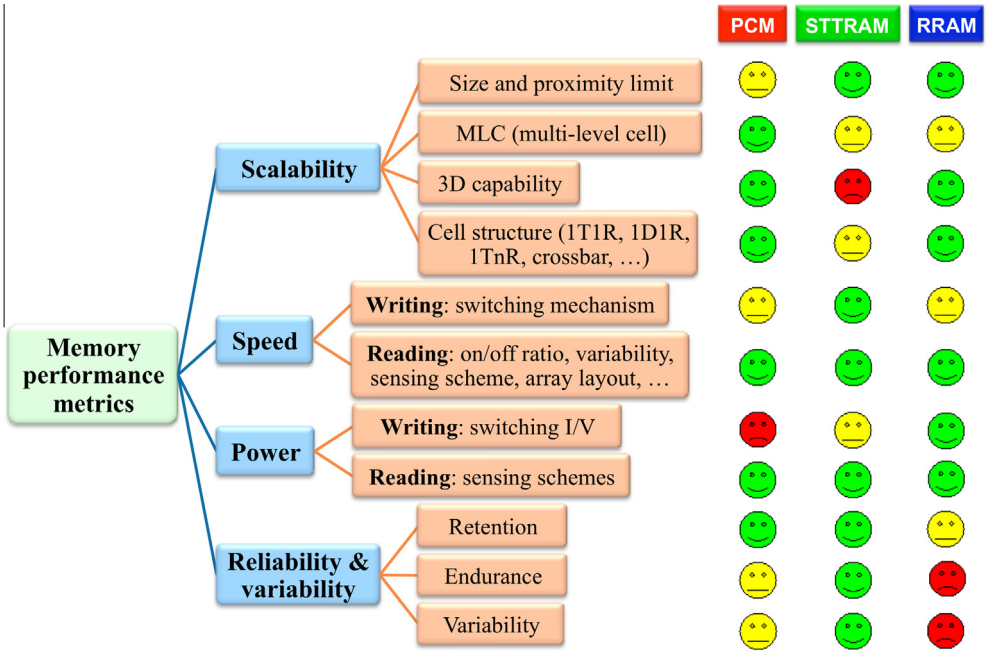

Fig. 5. Key metrics for memory performance assessment and a qualitative comparison of STTRAM, PCM, and RRAM based on the metrics.

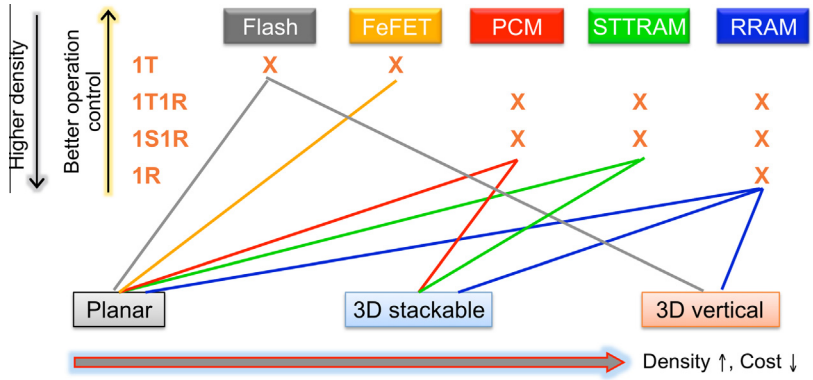

Fig. 6. Memory cell structure for major NVM devices: 1-transistor (1T), 1-transistor-1-resistor (1T1R), 1-selector-1-resistor (1S1R), and 1-resistor (1R). Different cell structures can be utilized in memory architectures from planar arrays to 3D stackable and 3D vertical arrays.

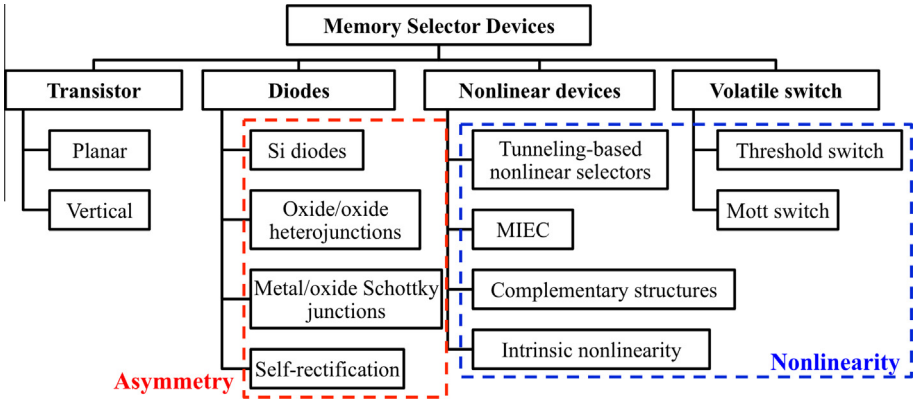

Fig. 7. Taxonomy of memory selector devices.

selected junctions, and therefore nonlinear selectors (with higher resistance at lower voltage) can significantly increase the effective resistance of unselected cells to reduce sneak leakage. In addition, the sneak paths in CBAs always include a segment of reverse current flow, and therefore asymmetric selectors (with higher resistance along reverse direction) can block leakage along sneak paths. Selectors need to provide sufficiently high on-current required for memory operation. Their on/off ratio determines feasible CBA size. Bipolar switching memories require bi-directional selectors (*e.g.*, nonlinear devices), while rectifying diodes only work for unipolar switching memories.

Fig. 8 summarizes the on-current density *vs.* rectification ratio of rectifying diode selectors and the on-current density *vs.* nonlinearity ratio of nonlinear selectors reported in literature [42–60]. Diode selectors include Si p-n junctions, Schottky diodes formed between metals and semiconducting oxides, and hetero-junctions between two oxides. Nonlinear selectors are often based on tunneling conduction mechanisms. It should be emphasized that summary plots like these are only snapshots of selector characteristics at certain voltage (*e.g.*, 1 V/2 V), which cannot represent full functionality of selectors in CBAs. Selector functionality and CBA performance have to be analyzed in a comprehensive framework

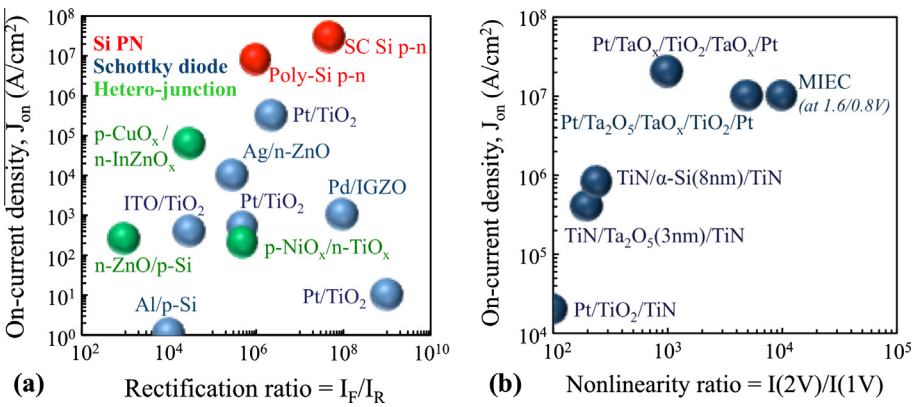

Fig. 8. Summary of on-current density *vs.* on/off ratio of (a) rectifying diode selectors [42–53] and (b) nonlinear selectors [54–60].

that incorporates voltage-dependent device characteristics and various array data patterns [40]. Both rectification ratio and non-linearity ratio are voltage dependent and should not be used as fixed parameters for selector assessment.

Fig. 9 summarizes the key parameters of volatile switch selectors, including metal-insulator-transition (MIT) switches and threshold switches [61–69]. These devices switch from an off-state with high resistance to an on-state with low resistance at a threshold voltage (V_{th}) and recover to the off-state when voltage drops below a hold voltage (V_{hold}). Their high on/off ratio could suppress sneak leakage in CBAs. Notice that many volatile switch selectors can achieve high on-current density, but some also suffer from relatively high off-leakage. The V_{th} and V_{hold} need to be compatible with the operation voltages of the memory elements, because applied voltage will redistribute between the selector and the memory element after the switching of either the selector or the memory element.

In addition to on/off ratio and on-current (density) as two basic requirements, selectors also need to have sufficient speed, scalability and endurance to avoid limiting performance of memory cells. Fig. 10 summarizes the cycling endurance, writing speed, and smallest size of some selectors reported in literature [42–69]. There has been clear improvement of selector properties over the years. Endurance over 10^{10} cycles, speed faster than 10 ns, and device size on the order of 10 nm have been demonstrated.

Two-terminal selectors have been explored mostly for PCM and RRAM, both of which are considered suitable for high-density low-cost storage applications. STTRAM, on the other hand, has targeted mostly high-performance applications that are less sensitive to cost. Since both working memory and data storage are needed in most computing systems, there would be a cost advantage to provide both solutions based on the same memory technology [70]. As STTRAM is the only emerging NVM with proven working memory performance, it would be useful if a high-density version of STTRAM can also be achieved. The small TMR of STTRAM limits the potential of MLC. The footprint of STTRAM is determined by the size of access transistors that has to be large enough to provide sufficient current for MTJ switching. While MTJ can be as small as several F^2 (F – feature size of a technology node), the size of access transistors is typically 50–100F^2 or even larger. If n MTJs can share one transistor ($n < A_{TR}/A_{MTJ}$, A_{TR}: transistor area, A_{MTJ}: MTJ area including spacing), effective bit density of STTRAM can be increased n times. However, a 1T-nMTJ array essentially forms a CBA with sneak leakage that has to be suppressed by two-terminal selectors. Selectors for MTJ would have more stringent requirements than those for RRAM, because of the low resistance of MTJ (typically several kΩ in comparison to tens of kΩ–MΩ for RRAM) and the small TMR. Since selectors divide voltage from MTJs, selector resistance needs to be comparable with MTJ resistance in the operation voltage range. An analysis has shown that with suitable bi-directional selectors, it may be possible to build functional 1T-nMTJ STTRAM for high-density applications [71]. If successful, a high-density STTRAM array enabled by proper selectors may be integrated with a high-performance STTRAM array to fulfill both working memory and data storage functions. This 1T-nMTJ structure may also provide an alternative for STTRAM design that has so far focused heavily on the reduction of MTJ writing current. In comparison with conventional STTRAM, VCMA-based STTRAM may be more suitable for this 1T-nMTJ design, due to its larger MTJ resistance (compatible with a wider range of selectors) and the possibility of unipolar switching (compatible with rectifying diode selectors).

4. Major NVM test-chips and performance

4.1. Summary of major NVM test-chips

Fig. 11 summarizes major industry test chips of PCM [72–83], STTRAM [84–98], and RRAM [99–113] reported in conferences (e.g., International Solid-State Circuits Conference, ISSCC). They are fully functional test chips with peripheral circuitries. Size of the symbols is proportional to the technology node of the CMOS process used for the fabrication. The company name, array capacity, and technology node for every demonstrated chip are labeled near each symbol. Among the three NVM technologies, PCM test chips (blue symbols) were demonstrated the earliest, and vary from tens of Mb to several Gb in capacity. Capacity of STTRAM test chips (red symbols) is typically smaller, because STTRAM targets performance-driven applications instead of high-density data storage. RRAM (green symbols) varies in the widest range in terms of test chip capacity and technology nodes, because different companies are targeting different applications for this technology, *e.g.*, eNVM for micro-controller units (MCUs), standalone data storage, *etc.*

Fig. 12 compares the capacity and memory cell footprint (in the unit of F^2) of selected test chips from Fig. 11. In general, test chips with larger capacity require smaller memory cell footprint. Most test chips use MOSFET as the selector; however, for memory cells smaller than 10F^2, highly scaled selectors are often needed, which may include vertical transistor selectors, two-terminal selectors, or even self-selecting memory cells. These extremely small memory cells are typically PCM and RRAM for data storage applications.

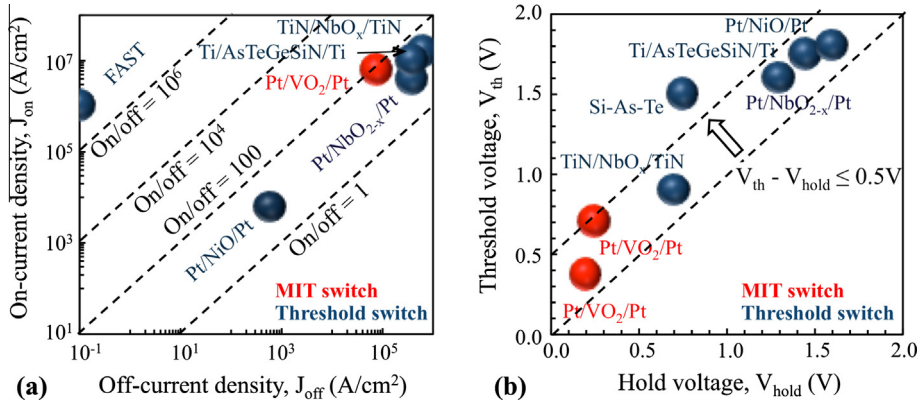

Fig. 9. Summary of reported volatile switch selector parameters: (a) on-current density *vs.* off-current density; (b) threshold voltage *vs.* hold voltage [61–69].

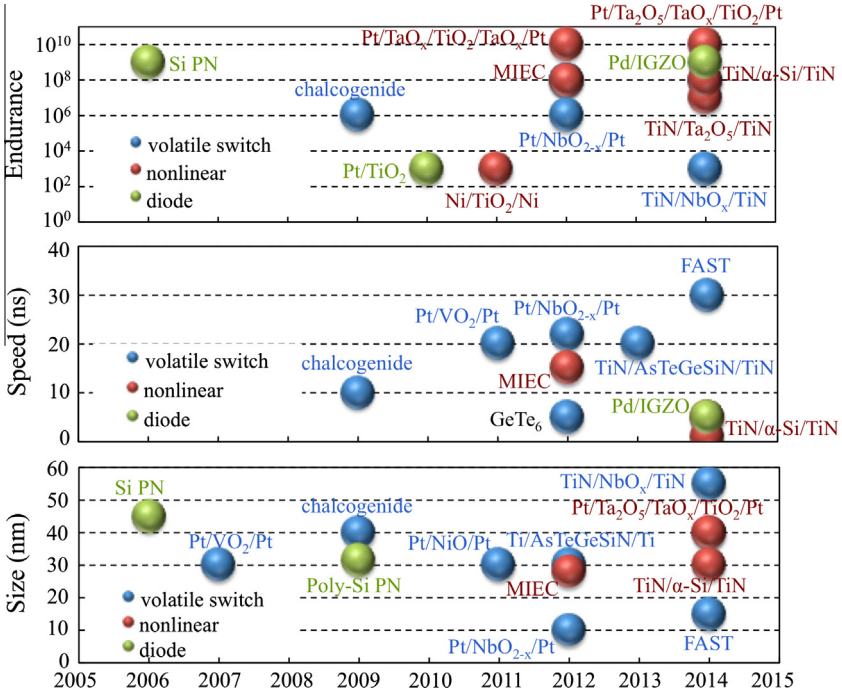

Fig. 10. Summary of cycling endurance, writing speed, and smallest device size of some reported selector devices, including rectifying diodes, nonlinear selectors, and volatile switch selectors [42–69].

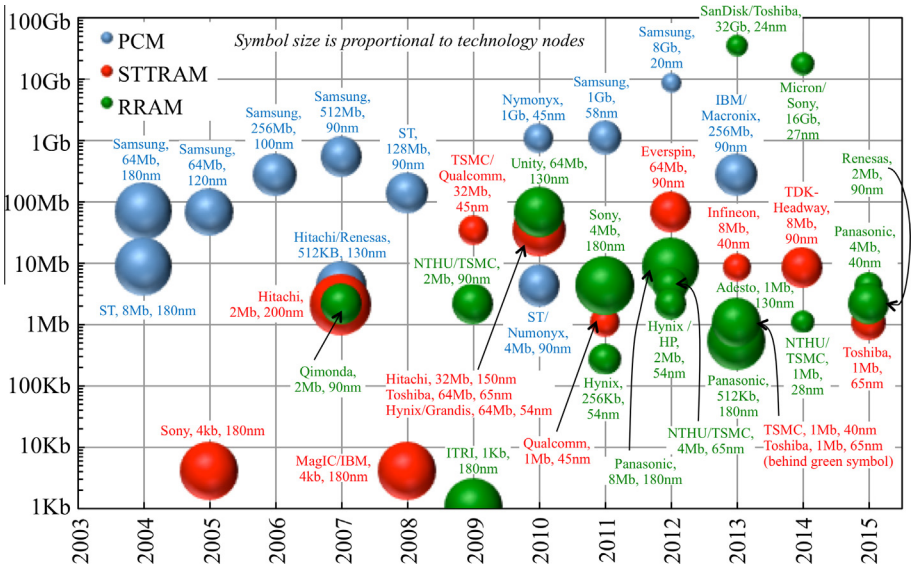

Fig. 11. Summary of industry test chips of PCM (blue symbols), STTRAM (red symbols), and RRAM (green symbols) [72–113], plotted as the total chip capacity *vs.* years. Symbol size is proportional to the technology node used for the fabrication of the test chips. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

4.2. NVM performance comparison

The performance of PCM, STTRAM, and RRAM is compared based on the reported parameters of these test chips, as shown in Fig. 13 where the *Y*-axis is the write speed and the typical cycling endurance of each type of NVMs is labeled in the figure. The symbol size is proportional to the memory cell footprint (F^2).

STTRAM demonstrates typical speed below 100 ns and sometimes even below 10 ns, which requires high enough switching current (and therefore large cell size). The measured endurance is in the range of 10^9–10^{12} cycles, where the upper bound is limited by the measurement time instead of device capability. PCM write speed varies from 100 ns up to 10 μs and its endurance is in the range of 10^6–10^9 cycles. RRAM test chip performance varies in a wider range, *e.g.*, write speed from as short as 10 ns up to over 100 μs and endurance in the range of 10^5–10^9 cycles. The wider range of RRAM performance reflects its less mature status as well as its larger variability. A test chip design has to accommodate the worst-performance devices. Both PCM and RRAM cells tend to be much smaller than STTRAM cells, due to lower switching current and more scalable selector devices.

Low power has become a key requirement for both logic and memory devices. NVM has zero standby power because voltage can be completely turned off without losing data. The active power

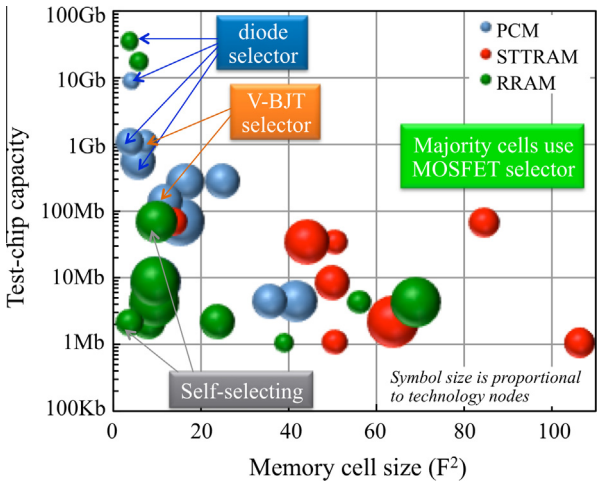

Fig. 12. Test chip capacity *vs.* memory cell size of selected test chips.

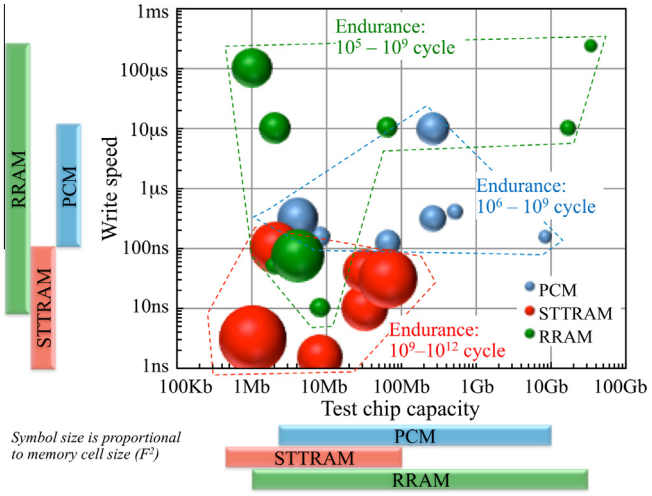

Fig. 13. Summary of emerging NVM parameters (chip capacity, cell footprint, write speed, and cycling endurance) based on reported measurement results of selected test chips from Fig. 11. Symbol size is proportional to memory cell footprint. The color bars near the *X*- and *Y*-axis indicate typical range of these NVM parameters.

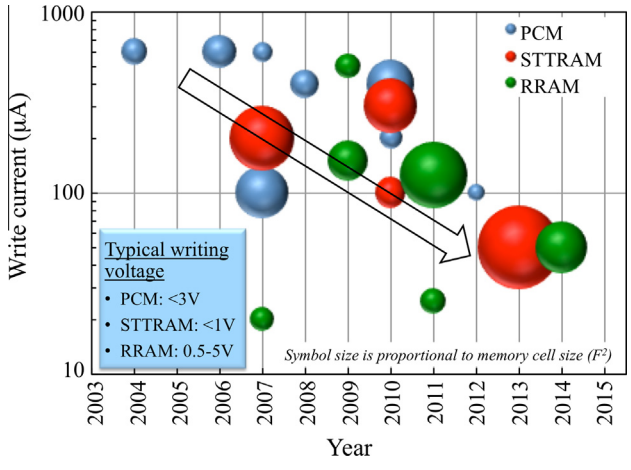

Fig. 14. The evolving trend of write current of major test chips, showing decreasing current (power) over time for all three major emerging NVMs. RRAM appears to have an advantage of ultra-low power.

is determined by write current and voltage. Although PCM, STTRAM, and RRAM have different write voltages, they are typically within 0.5–5 V. Therefore, the write current is the key factor that determines the write power, which is summarized in Fig. 14. Over the years, there is a trend of decreasing write current for all three NVMs. RRAM appears to have the greatest potential to achieve ultra-low write current and power, which is considered an advantage of RRAM in addition to its simplicity and low cost.

5. Emerging NVM applications

Emerging memories have a wide range of applications both in and beyond the memory space, as summarized in Table 2.

Fig. 15 shows the current memory hierarchy with typical access speed, as well as the performance range of STTRAM, FeFET, PCM, and RRAM. The speed lineup indicates the basic feasibility of each emerging memory for applications in the existing memory hierarchy; however, the adoption of a new memory technology involves many more complex factors.

5.1. Replace existing memories

Replacing existing memories is often the first justification for a new memory device, *e.g.*, STTRAM to replace SRAM (L2/L3 cache) or DRAM, RRAM to replace Flash memory, *etc.* Replacement is a straightforward objective with easily quantified benchmark criteria. However, existing technology, product requirements, and business infrastructure also impose high entry barriers for new technologies. Overcoming such barriers would require new technologies to provide significant advantages in performance, scalability, and cost.

Although STTRAM can reach the speed and endurance comparable to SRAM, its initial application may target lower performance requirements (*e.g.*, eFlash replacement) to ease the entry to production. Further process improvement and cost reduction may help to exploit the full potential of STTRAM. The performance of different RRAM devices varies in a wide range, but their primary advantages lie in scalability and simple structures/processing, suitable for high-density data storage. However, it is increasingly difficult to compete with vertical 3D NAND in terms of density and bit-cost. More promising opportunities for RRAM may exist in applications requiring the performance beyond the capabilities of Flash memories, *e.g.*, lower write voltage, faster speed, longer endurance, *etc.* RRAM has also been experimented for flexible electronics.

Embedded NVMs (eNVMs) play increasingly important functions in micro-controller units (MCUs) for industry and automotive applications, as well as consumer electronics [114,115]. The real-time code-execution/customization and data management capabilities enabled by eNVM improve system performance, enhance

Table 2
Summary of emerging NVM applications and challenges.

Applications	Key challenges
NVM application in the memory hierarchy	
• Replacing incumbent memories	• Cheaper and/or better
• Improving memory hierarchy (*e.g.*, SCM)	• Reliability; compatibility with existing infrastructure
NVM for low-power computing	
• Nonvolatile logic	• Match logic performance
• Fine-grained power gating (*e.g.*, "normally off" design)	• Application dependence
NVM applications beyond memory space	
• Synaptic devices for bio-inspired computing	• Controllable analog behaviors
• Hardware security primitives	• Reliability

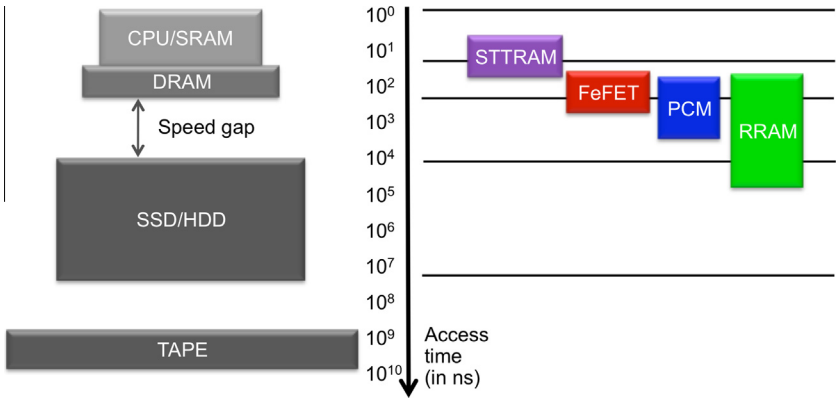

Fig. 15. An illustration of memory hierarchy with typical access speed, lined up with the performance range of major emerging NVM devices.

security, and lower cost. These have so far been fulfilled by one-time-programmable (OTP) memory, multiple-time-programmable (MTP) memory, and embedded Flash (*e.g.*, 1T NOR, split-gate flash). Embedded NVM needs to be compatible with logic platform, which becomes more challenging for existing solutions at advanced nodes and may create opportunities for emerging NVMs that are typically integrated in BEOL. Emerging NVMs can provide higher performance beyond the capability of eFlash. At the same time, reliability of emerging NVMs has to be further improved for embedded applications, especially in automotive space with stringent thermal stability and zero defect rate requirements. These BEOL memory devices have to withstand the thermal budget of logic process (typically 400 °C for one hour). Additional masks and processing steps need to be minimized to reduce cost. New materials and contamination have to be carefully controlled. Optimization for eNVM is different from that for standalone memories, in terms of robustness, stability, power efficiency, and cell size; therefore, different development strategies need to be adopted. Emerging NVM candidates offer a wide range of performance and scalability for different types of embedded applications.

5.2. Simplify memory-storage hierarchy and enable storage-class memory (SCM)

From high-speed L1 cache to high-density hard disk drive (HDD), access speed varies more than six to seven orders of magnitude. The memory hierarchy is designed to optimize the performance-cost tradeoff in existing memories. SRAM-based cache has accounted for a major portion of power consumption in mobile CPU or SoC. Scalable high-performance NVMs (*e.g.*, STTRAM) with the speed and endurance close to those of SRAM/DRAM can enable nonvolatile working memory and improve system performance. However, finding a "universal memory" with both high performance and low cost for a uniform memory-storage hierarchy is very challenging. The large variety of emerging NVMs also shows that the performance-cost tradeoff exists not only in incumbent memories but also among emerging NVMs.

While DRAM is slower than cache, it is several orders of magnitude faster than disk drive [116], as shown in Fig. 15. Closing this speed gap between memory and storage would have great impact on system performance and cost. Storage class memory (SCM) is introduced to fill this performance gap by reaching access speed close to DRAM and at the same time lowering the bit-cost to the level of HDD [117]. It combines the benefits of random access of memories and the archival capabilities of HDD. SCM requires a solid-state NVM with excellent scalability and preferably MLC capability, as well as extremely effective manufacturing at ultra-high areal density. SCM would greatly simplify the conventional memory and storage hierarchy. So far, PCM and RRAM have been considered the top device candidates for SCM.

5.3. NVM-enabled low-power computing solutions and novel architectures

Standby power has become the major source of power consumption in many computing systems. Emerging NVMs could eliminate standby power by turning off system power without losing data. Wireless sensor nodes and network for Internet of Things (IoT) has more restrictive power budget than mobile devices. These systems spend much of their lifetime in sleep mode and only wake up occasionally when triggered by certain signals. Volatile SRAM in short standby state still has to be kept in on-state to avoid data loss. SRAM can be turned off in long standby mode, but only after all the data have been transferred to permanent storage which takes long time (up to a few hundred μs). Low-voltage SRAM design and operation have limitation due to robustness degradation, and therefore larger cell design (*e.g.*, ten-transistor SRAM) may be required. Flash memories are not suitable for these applications because of their high writing voltage and power. Emerging NVMs with lower switching power/voltage offer opportunities to design "normally-off" memory systems to eliminate standby power. Nonvolatile lower-level cache would significantly accelerate data storage (*i.e.*, data transfer from register files to NV cache instead of solid-state drives) and enable more fine-grained power gating design. STTRAM is considered a promising candidate for this application owing to its high performance. On the other hand, STTRAM also consumes high power during switching; therefore, the benefit of saving standby power needs to outweigh the increase of active power, which depends on both activity factors in memory hierarchy and STTRAM device characteristics [118]. Reducing the active power of STTRAM through material and device engineering will eventually determine the feasibility of STTRAM for these low-power designs.

Data movement in von Neumann architectures consumes high power and slows down system speed. New architectures with localized data (at or near computing units) may achieve better speed and efficiency. Emerging NVMs offers promising technology options for non-von-Neumann architectures where logic and memory functions are more closely integrated or inter-mixed. By inserting simple two-terminal NVM elements (*e.g.*, MTJ, RRAM) in logic gates (*e.g.*, SRAM, adder, look-up-table), it is possible to create logic components with built-in memory to enable non-volatile information processing (NVIP), as shown in Fig. 16. Some emerging logic devices have built-in memory functions (*e.g.*, spintronic devices based on ferromagnetic polarization), which are also suitable NVIP components. NVIP not only reduces data movement to

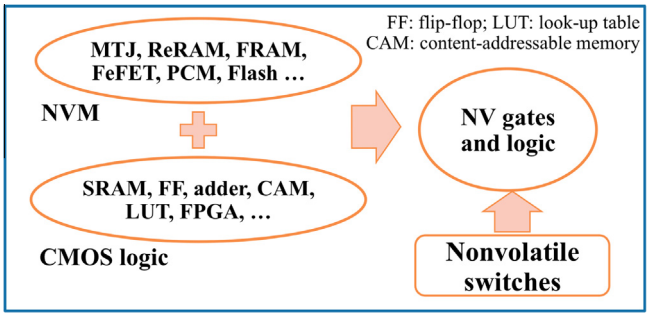

Fig. 16. Illustration of memory-logic integration in low-level circuit components for non-volatile information processing (NVIP).

increase throughput and lower power, but also enables novel architecture and designs, *e.g.*, latch-less pipeline design, systolic architectures. Even though the memory-related operation on these NVM elements (*e.g.*, store and recall of status data) may be less frequent than logic operation, the endurance and speed of these NVM elements have to be sufficiently good to avoid becoming performance bottleneck in these computing systems.

FeFET RAM resembles a so-called "negative-capacitance" field-effect-transistor (NC-FET), a type of steep sub-threshold slope logic device [119]. NC-FET utilizes ferroelectric gate dielectrics to trigger an internal amplification to achieve sub-60 mV/dec slope during transistor switching. Both FeFET and NC-FET can use doped ferroelectric HfO_x as a CMOS-compatible high-κ gate dielectric, although they are designed and operated differently. We can envision a "ferroelectric integrated circuit solution (FICS)" by integrating NC-FET for low-power logic and FeFET for non-volatile memory, which may create an ultra-low power computing system.

6. Novel functionalities beyond memory space

Some unique characteristics of emerging NVMs may be utilized for applications beyond memory space. Two examples are discussed here – neuromorphic computing and hardware security.

6.1. Brain-inspired computing and synaptic functions

Brain achieves very high performance with extremely low power in certain functions (*e.g.*, recognition, inference, *etc.*) in comparison to today's computers. In neural networks, synapses connect neurons and play key functions in the learning process. One of the well-known learning rules is the "spike time dependent plasticity" (STDP), *i.e.*, synaptic weight modified by the timing difference between pre- and post-synapses neuron signals [120]. It has been shown that the analog behavior of some emerging NVMs (*i.e.*, gradual resistance change controlled by pulse length or amplitude) can be utilized to imitate synaptic behaviors [121]. By fine-tuning the incremental change of NVM resistance with switching pulses, potentiation and depression of "synaptic weight" (typically represented by NVM resistance) can be implemented. Although using solid-state devices to mimic synapses is not new (*e.g.*, floating gate devices have been experimented for synapses), emerging NVMs may achieve the synaptic density close to that in brains ($\sim 10^{10}$ cm^{-2}), owing to their good scalability and low power. There have been many reports of device-level implementation of synaptic functions using both PCM and RRAM; however, array-level demonstration of neural network and algorithms is still limited [122]. The stability and precision of resistance modulation of NVM devices need to be further improved for large-scale array demonstrations. Gradual resistance change is sometimes difficult to achieve in certain switching operations of NVM devices (*e.g.*, set in RRAM, reset in PCM). Alternative designs have been proposed to overcome these limitations, *e.g.*, a "2-PCM synapse" to utilize gradual resistance decrease (set) and avoid the abrupt resistance increase (reset) in PCM [123]. The impact of imperfection of NVM devices on the performance of synaptic network is an important research topic [124].

6.2. Hardware security

Some stochastic behaviors of emerging NVMs are undesirable for memory applications, but could be utilized as entropy sources for security applications that embrace truly random variations. These behaviors include variability of NVM device parameters (*e.g.*, resistance, switching voltage, *etc.*), noise in read signal (*e.g.*, random telegraph noise, RTN), and probabilistic switching (*i.e.*, switching yield controlled by operation conditions). For example, RRAM switching is intrinsically stochastic with large variation in cell resistance and switching voltages. RTN is common in RRAM, due to charge capture and emission at the trap sites near the nano-gaps and filaments. Condition-dependent switching probability has been characterized for RRAM [125] and STTRAM [126]. These random behaviors of emerging NVMs can be utilized to generate hardware security primitives.

6.3. True random number generator (TRNG)

The probabilistic switching of STTRAM and RRAM allows programing of these devices with controlled probability by tuning programming pulse duration or amplitude. Controlled switching with a probability of 50% leads to equal chance of a device falling in "0" or "1" afterwards. This behavior can be utilized to create a TRNG, which has been demonstrated experimentally with the proof of true randomness for security applications [126–128]. Alternatively, the strong signal of RTN in RRAM can also be used to generate random numbers in a simple circuit [129].

6.4. Physical unclonable function (PUF)

PUF utilizes the physical randomness to generate unclonable instance-specific security features and can be used as "fingerprint" to identify or authenticate specific hardwares [130]. Existing PUF solutions mostly exploit inevitable IC manufacturing variation, which is uncontrollable, unclonable, and unique to each individual IC. The variability of RRAM provides an alternative source of randomness for PUF implementation. Unlike manufacturing variation that is fixed once an IC is fabricated, RRAM variability is intrinsic in physical mechanisms, which is less process dependent and potentially reconfigurable [131–133]. Fig. 17(a) shows an example of a PUF based on cell-to-cell resistance variation in a 1T1R RRAM array. The "challenges" of the PUF are the addresses of two n-bit data and the bit-wise comparison generates n-bit "responses", which form PUF challenge-response pairs (CRPs). The uncontrolled cell-to-cell variation is the source of randomness, unclonability, and uniqueness required for this PUF design.

PUF performance can be measured with a set of metrics. Among them, three key parameters are bias (to measure uniformity of bit 0 or 1 in the response), inter-chip Hamming Distance (to measure the uniqueness of PUFs), and intra-chip Hamming Distance (to measure the reliability of PUFs). Fig. 17(b) illustrates how the three parameters are defined along different measurement dimensions to evaluate PUFs. An ideal PUF should generate responses with equal percentage of bit 0 s and 1 s, *i.e.*, bias value of 50%. Different responses from different PUFs for the same challenge should be completely different and uncorrelated, *i.e.*, inter-chip HD of 50%. Finally, the same PUF should always generate the same response

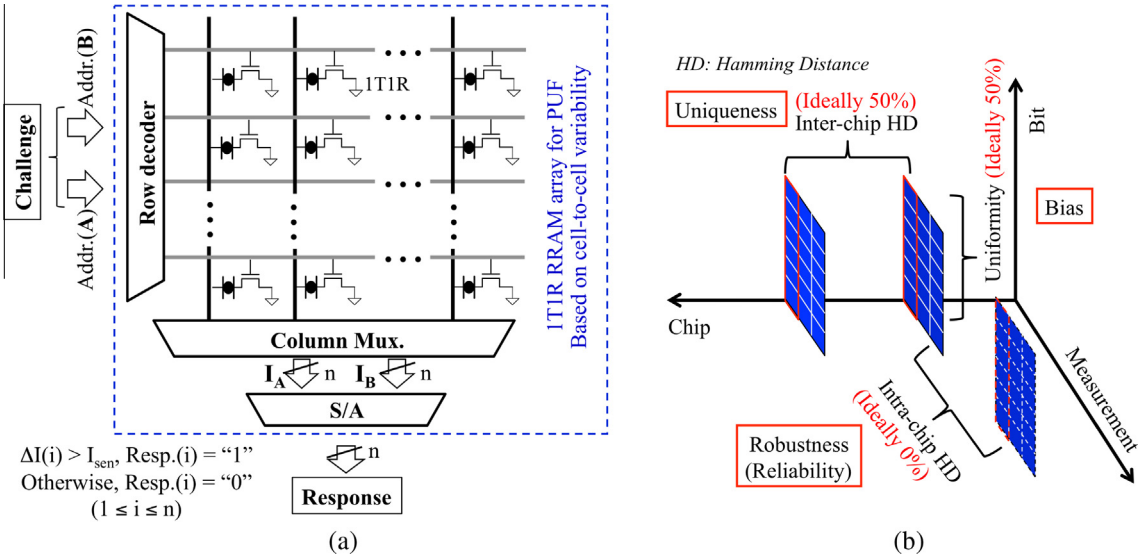

Fig. 17. (a) Illustration of a PUF implementation based on cell-to-cell variation in a 1T1R RRAM array [132]; (b) illustration of key PUF parameters: bias, uniqueness, and relibility.

when inquired by the same challenge at different time and measurement conditions, *i.e.*, intra-chip HD of 0%.

The reliability (intra-chip HD) of RRAM-based PUF is strongly affected by the non-ideal behaviors of RRAM, including reading instability (*e.g.*, due to RTN), thermal dependence of RRAM resistance, and retention loss [132]. Fig. 18a shows simulated inter-chip and intra-chip HDs with all these behaviors considered and three types of retention loss. The inter-chip HD is not affected because the randomness in RRAM resistance variation is preserved in these behavior models, but intra-chip HD significantly increases, especially with retention loss over time. Fig. 18b calculates the false acceptance rate (FAR) and false rejection rate (FRR) of RRAM PUF used in an identification application, assuming binominal distributions for both inter-chip and intra-chip HDs. The identification capability of RRAM PUF degrades over time due to retention loss (curve shifting to the upper right corner).

PUF has also been implemented in other emerging NVMs, including PCM [134] and STTRAM [135]. NVM-based PUFs are much smaller than other PUFs (*e.g.*, SRAM PUF); therefore, they not only are more suitable for lightweight security applications but also allow longer bit-length to enhance security solutions. The re-programmability of NVMs can also be utilized in creative PUF designs to improve reliability [132]. Table 3 summarizes the requirements on emerging NVM devices for memory and PUF applications, indicating different strategies required for technology optimization in these applications.

Table 3
Summary of NVM requirements for memory and PUF applications.

Property	For memory	For PUF
Speed	Important	Useful, not critical
Power/energy	Important	Useful
Cycling endurance	Important	Not important
Retention	Important	Important
Reading stability	Important	Important
Variability	Undesirable	Important
Forming	Undesirable	Useful
Scalability	Important	Useful

7. Summary

Emerging NVMs offers a wide range of performance, maturity, and scaling potential. PCM, STTRAM, RRAM, and FeFET memory stand out as the most promising emerging NVM candidates. Emerging NVMs provide opportunities to improve or replace incumbent memories, simplify memory-storage hierarchy, design

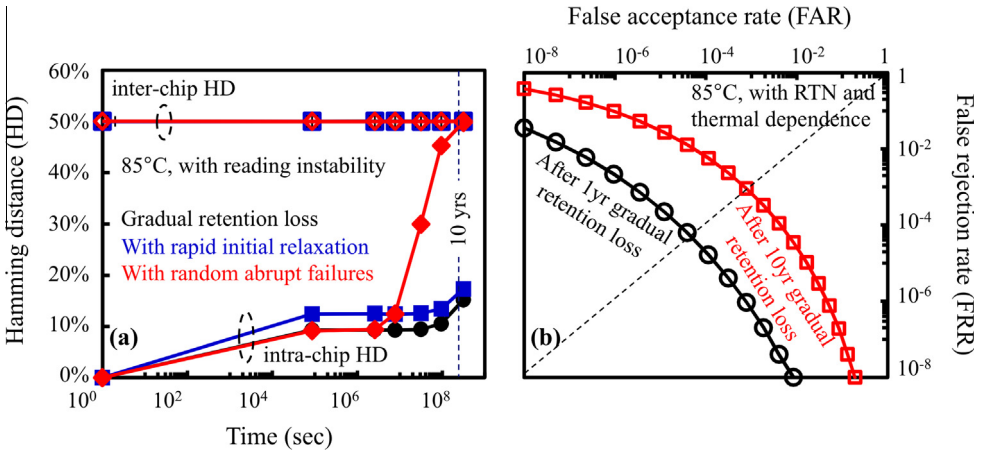

Fig. 18. (a) Calculated RRAM PUF inter-chip and intra-chip HDs at 85 °C operation, with thermal dependence and retention loss included; (b) calculated error rates of an identification application using RRAM PUF [88].

novel architectures, and enable new functionalities. Scalable selectors play an important role in memory device and architecture design. There are still many challenges in high-yield NVM manufacturing, material and device engineering, and the optimization of suitable emerging NVM solutions in different applications. However, significant progress has been made on all major emerging NVM candidates, and some of them have entered early production. Future electronics driven by mobile, IoT, and data-centric applications may rely more on NVM technologies to meet the demand for high-density data storage, ultra-low power design, intelligent hardware, and even high-performance computing systems.

References

[1] Wulf WA, McKee SA. Hitting the memory wall: implications of the obvious. ACM SIGARCH Comp Arch News 1995;23(March):20–4.
[2] ITRS Emerging Research Devices Chapter, 2013.
[3] Kreupl F et al. Carbon-based resistive memory. IEDM Tech Dig December 2008.
[4] Li Y, Sinitskii A, Tour JM. Electronic two-terminal bistable graphitic memories. Nat Mater 2008;7(12):966–71.
[5] Hwang SK et al. Flexible multilevel resistive memory with controlled charge trap band N-doped carbon nanotubes. Nano Lett 2012;12(5):2217–21.
[6] Song H, Reed MA, Lee T. Single molecule electronic devices. Adv Mater 2011;23(14):1583–608.
[7] Pro T et al. Investigation of hybrid molecular/silicon memories with redox-active molecules acting as storage media. IEEE Trans Nanotechnol 2009;8 (2):204–13.
[8] McWilliams CR, Celinska J, Paz de Araujo CA, Xue K-H. Device characterization of correlated electron random access memories. J Appl Phys 2011;109(9). p. 091608-1–6.
[9] Ha SD, Aydogdu GH, Ramanathan S. Metal-insulator transition and electrically driven memristive characteristics of $SmNiO_3$ thin films. Appl Phys Lett 2011;98. p. 012105-1–3.
[10] Stoliar P et al. Universal electric-field-driven resistive transition in narrow-gap Mott insulators. Adv Mater 2013;25(23):3222–6.
[11] Liu S-H et al. High-performance polyimide-based ReRAM for nonvolatile memory application. IEEE Electron Dev Lett 2013;34(1):123–5.
[12] Kim JJ, Cho B, Kim KS, Lee T, Jung GY. Electrical characterization of unipolar organic resistive memory devices scaled down by a direct metal-transfer method. Adv Mater 2011;23(18):2104–7.
[13] Burr GW et al. Phase change memory technology. J Vac Sci Technol, B 2010;28:223–62.
[14] Raoux S, Xiong F, Wuttig M, Pop E. Phase change materials and phase change memory. MRS Bull 2014;39:703–10.
[15] Lai S, Lowrey T. OUM – A 180 nm nonvolatile memory cell element technology for stand alone and embedded applications. IEDM Tech Dig December 2001. p. 36.5.1–36.5.4.
[16] Lam CH. Phase change memory and its intended applications. IEDM Tech Dig December 2014:689–92.
[17] Cheng HY et al. Atomic-level engineering of phase change material for novel-switching and high-endurance PCM for storage class memory application. IEDM Tech Dig December 2013:758–61.
[18] Slaugher JM et al. High density ST-MRAM technology. IEDM Tech Dig December 2012:673–6.
[19] Ikeda S et al. A perpendicular-anisotropy CoFeB–MgO magnetic tunnel junction. Nat Mater 2010;9:721–4.
[20] Khvalkovskiy AV et al. Basic principles for STT-MRAM cell operation in memory arrays. J Phys D Appl Phys 2013;46. 074001120.
[21] Jan G et al. Demonstration of fully functional 8Mb perpendicular STT-MRAM chips with sub-5ns writing for non-volatile embedded memories. In: Symp VLSI Tech. p. 50–1.
[22] Maruyama T et al. Large voltage-induced magnetic anisotropy change in a few atomic layers of iron. Nat Nanotechnol 2009;4(January):158–61.
[23] Liu L et al. Spin-torque switching with the giant spin Hall effect of tantalum. Science 2012;336(May):555–8.
[24] Baek IG et al. Highly scalable non-volatile resistive memory using simple binary oxide driven by asymmetric unipolar voltage pulses. IEDM Tech Dig December 2004.
[25] Lee HY et al. Low power and high speed bipolar switching with a thin reactive Ti buffer layer in robust HfO_2 based RRAM. IEDM Tech Dig December 2008.
[26] Wei Z et al. Highly reliable TaOx ReRAM and direct evidence of redox reaction mechanism. IEDM Tech Dig December 2008.
[27] Kim MJ et al. Low power operating bipolar TMO ReRAM for sub 10 nm era. IEDM Tech Dig December 2010:444–7.
[28] Liu TY, et al. A 130.7 mm^2 2-layer 32Gb ReRAM memory device in 24nm technology. ISSCC, 12.1, February 2013.
[29] Valov I, Waser R, Jameson JR, Kozicki M. Electrochemical metallization memories–fundamentals, applications, prospects. Nanotechnology 2011;22 (25). 254003122.
[30] Fackenthal R, et al. A 16Gb ReRAM with 200MB/s write and 1GB/s read in 27 nm technology. ISSCC, 19.7, February 2014.
[31] Chen YY et al. Understanding of the endurance failure in scaled HfO_2-based 1T1R RRAM through vacancy mobility degradation. IEDM Tech Dig December 2012:482–5.
[32] Fantini A et al. Intrinsic program instability in HfO_2 RRAM and consequences on program algorithms. IEDM Tech Dig December 2015:169–72.
[33] Fantini A et al. Intrinsic switching variability in HfO_2 RRAM. Inter Memory Workshop 2013(May):30–3.
[34] Sune J. New physics-based analytic approach to the thin-oxide break- down statistics. IEEE Electron Dev Lett 2001;22(6):296–8.
[35] Bersuker G et al. Metal oxide resistive memory switching mechanism based on conductive filament properties. J Appl Phys 2011;110. p. 124518-1–7.
[36] Ma TP, Han J-P. Why is nonvolatile ferroelectric memory field-effect transistor still elusive? IEEE Electron Dev Lett 2002;23:386–8.
[37] Muller J et al. Ferroelectric hafnium oxide: a CMOS-compatible and highly scalable approach to future ferroelectric memories. IEDM Tech Dig December 2013:280–3.
[38] Tsymbal EY, Kohlstedt H. Tunneling across a ferroelectric. Science 2006;313 (5784):181–3.
[39] Govoreanu B et al. Vacancy-modulated conductive oxide resistive RAM (VMCO-RRAM): an area-scalable switching current, self-compliant, highly nonlinear and wide on/off-window resistive switching cell. IEDM Tech Dig December 2013:256–9.
[40] Chen A. Comprehensive methodology for the design and assessment of crossbar memory array with nonlinear and asymmetric selector devices. IEDM Tech Dig December 2013:746–9.
[41] Chen A. Nonlinearity and asymmetry for device selection in cross-bar memory arrays. IEEE Trans Electron Dev 2015;62(September):2857–64.
[42] Sasago Y et al. Cross-point phase change memory with 4F2 cell size driven by low-contact-resistivity poly-Si diode. Symp VLSI Tech Dig June 2009:24–5.
[43] Oh JH et al. Full integration of highly manufacturable 512Mb PRAM based on 90 nm technology. IEDM Tech Dig December 2006:515–8.
[44] Huby N et al. New selector based on zinc oxide grown by low temperature atomic layer deposition for vertically stacked non-volatile memory devices. Microelectron Eng 2008;85(12):2442–4.
[45] Cho B et al. Rewritable switching of one diode–one resistor nonvolatile organic memory devices. Adv Mater 2010;22(11):1228–32.
[46] Park WY et al. A $Pt/TiO_2/Ti$ Schottky-type selection diode for alleviating the sneak current in resistance switching memory arrays. Nanotechnology 2010;21(19). p. 195201-1–4.
[47] Shin YC et al. $(In, Sn)_2O_3/TiO_2/Pt$ Schottky-type diode switch for the TiO_2 resistive switching memory array. Appl Phys Lett 2008;92(16). p. 162904-1–3.
[48] Kim GH et al. Schottky diode with excellent performance for large integration density of crossbar resistive memory. Appl Phys Lett 2012;100(21). p. 213508-1–3.
[49] Huang JJ, Kuo CW, Chang WC, Hou TH. Transition of stable rectification to resistive-switching in $Ti/TiO_2/Pt$ oxide diode. Appl Phys Lett 2010;96(26). p. 262901-1–3.
[50] Chasin A et al. High-performance a-IGZO thin film diode as selector for cross-point memory application. IEEE Electron Dev Lett 2014;35(6):642–4.
[51] Choi Y et al. High current fast switching n-ZnO/p-Si diode. J Phys D Appl Phys 2010;43. p. 345101-1–4.
[52] Ahn SE et al. Stackable all-oxide-based nonvolatile memory with Al_2O_3 antifuse and $p\text{-}CuO_x/n\text{-}InZnO_x$ diode. IEEE Electron Dev Lett 2009;30 (5):550–2.
[53] Lee MJ et al. A low-temperature-grown oxide diode as a new switch element for high-density, nonvolatile memories. Adv Mater 2007;19(1):73–6.
[54] Shin J et al. TiO_2-based metal-insulator-metal selection device for bipolar resistive random access memory cross-point application. J Appl Phys 2011;109(3). p. 033712-1–4.
[55] Lee W et al. Varistor-type bidirectional switch ($J_{MAX} > 107$ A/cm^2, selectivity $\sim 10^4$) for 3D bipolar resistive memory arrays. Symposium VLSI Tech June 2012:37–8.
[56] Govoreanu B, Adelmann C, Redolfi A, Zhang L, Clima S, Jurczak M. High-performance metal-insulator-metal tunnel diode selectors. IEEE Electron Dev Lett 2014;35(1):63–5.
[57] Zhang L et al. Ultrathin metal/amorphous-silicon/metal diode for bipolar RRAM selector applications. IEEE Electron Dev Lett 2014;35(2):199–201.
[58] Huang J et al. Bipolar nonlinear $Ni/TiO_2/Ni$ selector for 1S1R crossbar array applications. IEEE Electron Dev Lett 2011;32(10):1427–9.
[59] Virwani K et al. Sub-30 nm scaling and high-speed operation of fully-confined access-devices for 3D crosspoint memory based on mixed-ionic-electronic-conduction (MIEC) materials. IEDM Tech Dig December 2012:36–9.
[60] Woo J et al. Multi-layer tunnel barrier ($Ta_2O_5/TaO_x/TiO_2$) engineering for bipolar RRAM selector applications. VLSI Tech Sym June 2013:168–9.
[61] Kim S et al. Ultrathin (<10 nm) Nb_2O_5/NbO_2 hybrid memory with both memory and selector characteristics for high density 3D vertically stackable RRAM applications. Symposium VLSI Tech June 2012:155–6.
[62] Lee JH et al. Threshold switching in Si-As-Te thin film for the selector device of crossbar resistive memory. Appl Phys Lett 2012;100(12). p. 123505-1–4.
[63] Lee MJ et al. Highly-scalable threshold switching select device based on chalcogenide glasses for 3D nanoscaled memory arrays. IEDM Tech Dig December 2012:33–5.
[64] Lee M-J et al. A simple device unit consisting of all NiO storage and switch elements for multilevel terabit nonvolatile random access memory. ACS Appl Mater Interfaces 2011;3(11):4475–9.

[65] Lee M-J et al. Two series oxide resistors applicable to high speed and high density nonvolatile memory. Adv Mater 2007;19(22):3919–23.
[66] Son M et al. Excellent selector characteristics of nanoscale VO_2 for high-density bipolar ReRAM applications. IEEE Electron Dev Lett 2011;32(11):1579–81.
[67] Kau D et al. A stackable cross point phase change memory. IEDM Tech Dig December 2009:617–20.
[68] Jo SH et al. 3D-stackable crossbar resistive memory based on field assisted superlinear threshold (FAST) selector. IEDM Tech Dig December 2015:160–3.
[69] Kim WG et al. NbO_2-based low power and cost effective 1S1R switching for high density cross point ReRAM application. VLSI Tech Sym June 2014:138–9.
[70] Aitken R et al. Device and technology implications of the internet of things. Sym VLSI Tech June 2015:2–5.
[71] Chen A. Feasibility analysis of high-density spin-transfer-torque random-access-memory with shared access transistor structure. IEEE Electron Dev Lett 2015;36(December):1325–8.
[72] Cho WY, et al. A 0.18 μm 3.0 V 64Mb non-volatile phase-transition random-access memory (PRAM). ISSCC, 2.1, February 2004.
[73] Bedeschi F et al. An 8 Mb demonstrator for high-density 1.8 V phase-change memories. VLSI Circ Sym June 2004:442–5.
[74] Oh H, et al. Enhanced write performance of a 64Mb phase-change random access memory. ISSCC, 2.3, February 2005.
[75] Kang S, et al. A 0.1 μm 1.8 V 256Mb 66 MHz synchronous burst PRAM. ISSCC, 7.5, February 2006.
[76] Hanzawa S, et al. A512kB embedded phase change memory with 416kB/s write throughput at 100 μA cell write current. ISSCC 26.2, February 2007.
[77] Lee KJ, et al. A 90 nm 1.8 V 512 Mb diode-switch PRAM with 266MB/s read throughput. ISSCC, 26.1, February 2007.
[78] Bedeschi F, et al. A multi-level-cell bipolar-selected phase-change-memory. ISSCC, 23.5, February 2008.
[79] Villa C, et al. A 45 nm 1 Gb 1.8 V phase-change memory. ISSCC, 14.8, February 2010.
[80] Sandre GD, et al. A 90 nm 4 Mb embedded phase-change memory with 1.2 V 12 ns read access time and 1MB/s write throughput. ISSCC, 14.7, February 2010.
[81] Chung H, et al. A 58 nm 1.8 V 1 Gb PRAM with 6.4 MB/s program BW. ISSCC 28.7, February 2011.
[82] Choi Y, et al. A 20 nm 1.8 V 8 Gb PRAM with 40 MB/s program bandwidth. ISSCC, 2.5, February 2012.
[83] Close GF et al. A 256-Mcell phase-change memory chip operating at 2+ bit/cell. IEEE Trans Circ Syst 2013;60(June):1521–33.
[84] Hosomi M et al. A novel nonvolatile memory with spin torque transfer magnetization switching: spin-RAM. IEDM Tech Dig December 2005.
[85] Kawahara T, et al. 2Mb spin-transfer torque RAM (SPRAM) with bit-by-bit bidirectional current write and parallelizing-direction current read. ISSCC, 26.5, February 2007.
[86] Beach R et al. A statistical study of magnetic tunnel junctions for high-density spin torque transfer-MRAM (STT-MRAM). IEDM Tech Dig December 2008.
[87] Lin CJ et al. 45 nm low power CMOS logic compatible embedded STT MRAM utilizing a reverse-connection 1T/1MTJ cell. IEDM Tech Dig December 2009:279–82.
[88] Takemura R et al. A 32-Mb SPRAM with 2T1R memory cell, localized bi-directional write driver and '1'/'0' dual-array equalized reference scheme. IEEE J Solid-State Circ 2010;45(April):869–79.
[89] Tsuchida K, et al. A 64Mb MRAM with clamped-reference and adequate-reference schemes. ISSCC, 14.2, February 2010.
[90] Chung S, et al. Fully integrated 54nm STT-RAM with the smallest bit cell dimension for high density memory application; December 2010. p. 304–7.
[91] Kim JP et al. A 45 nm 1Mb embedded STT-MRAM with design techniques to minimize read-disturbance. VLSI Circ Sym June 2011:296–7.
[92] Slaughter JM et al. High density ST-MRAM technology. IEDM Tech Dig December 2012:673–6.
[93] Yu HC, et al. Cycling endurance optimization scheme for 1Mb STT-MRAM in 40 nm technology. ISSCC, 12.8, February 2013.
[94] Noguchi H et al. A 250-MHz 256b-I/O 1-Mb STT-MRAM with advanced perpendicular MTJ based dual cell for nonvolatile magnetic caches to reduce active power of processors. VLSI Circ Sym June 2013:108–9.
[95] Jefremow M, et al. Time-differential sense amplifier for sub-80mV bitline voltage embedded STT-MRAM in 40 nm CMOS. ISSCC, 12.4, February 2013.
[96] Jan G et al. Demonstration of fully functional 8 Mb perpendicular STT-MRAM chips with sub-5 ns writing for non-volatile embedded memories. VLSI Tech Sym June 2014:108–9.
[97] Noguchi H, et al. A 3.3 ns-access-time 71.2 μW/MHz 1Mb embedded STT-MRAM using physically eliminated read-disturb scheme and normally-off memory architecture. ISSCC, 7.5, February 2015.
[98] Kim C, et al. A covalent-bonded cross-coupled current-mode sense amplifier for STT-MRAM with 1T1MTJ common source-line structure array. ISSCC 7.4, February 2015.
[99] Dietrich S et al. A nonvolatile 2-Mbit CBRAM memory core featuring advanced read and program control. IEEE J Sold-State Circ 2007;42(April):839–45.
[100] Sheu SS et al. A 5 ns fast write multi-level non-volatile 1K bits RRAM memory with advance write scheme. VLSI Circ Sym June 2009:82–3.
[101] Tseng YH et al. High density and ultra small cell size of contact ReRAM (CR-RAM) in 90 nm CMOS logic technology and circuits. IEDM Tech Dig December 2009:109–12.
[102] Chevallier CJ, et al. A 0.13 μm 64 Mb multi-layered conductive metal- oxide memory. ISSCC, 14.3, February 2010.
[103] Sheu SS, et al. A 4 Mb embedded SLC resistive-RAM macro with 7.2 ns read-write random-access time and 160 ns MLC-access capability. ISSCC, 11.2, February 2011.
[104] Otsuka W, et al. A 4 Mb conductive-bridge resistive memory with 2.3 GB/s read-throughput and 216 MB/s program-throughput. ISSCC, 11.7, February 2011.
[105] Yi J et al. Highly reliable and fast nonvolatile hybrid switching ReRAM memory using thin Al_2O_3 demonstrated at 54 nm memory array. VLSI Tech Sym June 2011:48–9.
[106] Kawarara A, et al. An 8 Mb multi-layered cross-point ReRAM macro with 443 MB/s write throughput. ISSCC, 25.6, February 2012.
[107] Chang MF, et al. A 0.5 V 4 Mb logic-process compatible embedded resistive RAM (ReRAM) in 65 nm CMOS using low-voltage current-mode sensing scheme with 45 ns random read time. ISSCC, 25.7, February 2012.
[108] Lee HD et al. Integration of $4F^2$ selector-less crossbar array 2Mb ReRAM based on transition metal oxides for high density memory applications. IEEE Symp VLSI Technol June 2012:151–2.
[109] Kawahara A, et al. Filament scaling forming technique and level-verify-write scheme with endurance over 107 cycles in ReRAM. ISSCC 12.6, February 2013.
[110] Jameson JR et al. Conductive-bridge memory (CBRAM) with excellent high-temperature retention. IEDM Tech Dig December 2013:738–41.
[111] Chang MF, et al. Embedded 1 Mb ReRAM in 28 nm CMOS with 0.27-to-1V read using swing-sample-and-couple sense amplifier and self-boost-write-termination scheme. ISSCC, 19.4, February 2014.
[112] Hayakawa Y et al. Highly reliable TaO_x ReRAM with centralized filament for 28-nm embedded application. VLSI Tech Sym June 2015:14–5.
[113] Ueki M et al. Low-power embedded ReRAM technology for IoT applications. VLSI Tech Sym June 2015:108–9.
[114] Yamauchi T. Prospect of embedded non-volatile memory in the smart society. VLSI-TSA, April 2015.
[115] Baker K. Embedded nonvolatile memories. IMW, May 2012.
[116] Freitas RF, Wilcke WW. Storage-class memory: the next storage system technology. IBM J Res Dev 2008;52:439–47.
[117] Burr GW et al. Overview of candidate device technologies for storage-class memory. IBM J Res Dev 2008;52:449–64.
[118] Kitagawa E et al. Impact of ultra low power and fast write operation of advanced perpendicular MTJ on power reduction for high-performance mobile CPU. IEDM Tech Dig December 2012:677–80.
[119] Salahuddin S, Datta S. Use of negative capacitance to provide a sub-threshold slope lower than 60 mV/decade. Nano Lett 2008;8:405–10.
[120] Dan Y, Poo MM. Spike timing-dependent plasticity: from synapse to perception. Physiol Rev 2006;86(July):1033–48.
[121] Choi H et al. An electrically modifiable synapse array of resistive switching memory. Nanotechnology 2009;20. p. 345201-1–5.
[122] Jo SH et al. Nanoscale memristor device as synapse in neuromorphic systems. Nano Lett 2010;10(March):1297–301.
[123] Suri M et al. Phase change memory as synapse for ultra-dense neuromorphic systems: application to complex visual pattern extraction. IEDM Tech Dig December 2011:79–82.
[124] Burr G et al. Experimental demonstration and tolerancing of a large-scale neural network (165,000 synapses), using phase-change memory as the synaptic weight element. IEDM Tech Dig December 2014:697–700.
[125] Chen A, Lin M. Reset switching probability of resistive switching devices. IEEE Electron Dev Lett 2011;32(May):590–2.
[126] Choi WH et al. A magnetic tunnel junction based true random number generator with conditional perturb and real-time output probability tracking. IEDM Tech Dig December 2014:315–8.
[127] Fukushima A et al. Spin dice: a scalable truly random number generator based on spintronics. Appl Phys Exp 2014;7. p. 083001-1–4.
[128] Balatti S, Ambrogio S, Wang Z, Ielmini D. True random number generation by variability of resistive switching in oxide-based devices. IEEE J Emer Sel Top Circ Sys 2015;5(June):214–21.
[129] Huang CY et al. A contact-resistive random-access-memory-based true random number generator. IEEE Electron Dev Lett 2012;33(August):1108–10.
[130] Herder C et al. Physical unclonable functions and applications: a tutorial. Proc IEEE 2014;102(August):1126–41.
[131] Chen A. Utilizing the variability of resistive random access memory to implement reconfigurable physical unclonable functions. IEEE IEEE Electron Dev Lett 2015;36(February):138–40.
[132] Chen A et al. Comprehensive assessment of RRAM-based PUF for hardware security applications. IEDM Tech Dig December 2015:265–8.
[133] Liu R, Wu H, Pan Y, Qian H, Yu S. Experimental characterization of physical unclonable function based on 1 kb resistive random access memory arrays. IEEE Electron Dev Lett 2015;36(December):1380–3.
[134] Zhang L, Kong ZH, Chang C-H, Cabrini A, Torelli G. Exploiting process variations and programming sensitivity of phase change memory for reconfigurable physical unclonable functions. IEEE Trans Inf Foren Sec 2014;9(June):921–32.
[135] Zhang L, Fong X, Chang C-H, Kong ZH, Roy K. Highly reliable memory-based physical unclonable function using spin-transfer torque MRAM. IEEE ISCAS, June 2014. p. 2169–72.