

Resistive switching memory for high density storage and computing

Xiao-Xin Xu(许晓欣), Qing Luo(罗庆), Tian-Cheng Gong(龚天成), Hang-Bing Lv(吕杭炳), Qi Liu(刘琦), Ming Liu(刘明)

Citation: Chin. Phys. B, 2021, 30 (**5**): 058702. DOI: 10.1088/1674-1056/abe0c4
Journal homepage: http://cpb.iphy.ac.cn; http://iopscience.iop.org/cpb

What follows is a list of articles you may be interested in

Oscillation of S5 helix under different temperatures in determination of the open probability of TRPV1 channel

Tie Li(李铁), Jun-Wei Li(李军委), Chun-Li Pang(庞春丽), Hailong An(安海龙), Yi-Zhao Geng(耿轶钊), Jing-Qin Wang(王景芹)

Chin. Phys. B, 2020, 29 (**9**): 098701. DOI: 10.1088/1674-1056/aba600

Improvements in continuum modeling for biomolecular systems

Yu Qiao(乔瑜) and Ben-Zhuo Lu(卢本卓)

Chin. Phys. B, 2016, 25 (**1**): 018705. DOI: 10.1088/1674-1056/25/1/018705

Cosine fitting radiography and computed tomography

Li Pan-Yun, Zhang Kai, Huang Wan-Xia, Yuan Qing-Xi, Wang Yan, Ju Zai-Qiang, Wu Zi-Yu, Zhu Pei-Ping

Chin. Phys. B, 2015, 24 (**6**): 068704. DOI: 10.1088/1674-1056/24/6/068704

Fast parallel algorithm for three-dimensional distance-driven model in iterative computed tomography reconstruction

Chen Jian-Lin, Li Lei, Wang Lin-Yuan, Cai Ai-Long, Xi Xiao-Qi, Zhang Han-Ming, Li Jian-Xin, Yan Bin

Chin. Phys. B, 2015, 24 (**2**): 028703. DOI: 10.1088/1674-1056/24/2/028703

TOPICAL REVIEW — Physics in neuromorphic devices

Resistive switching memory for high density storage and computing*

Xiao-Xin Xu(许晓欣), Qing Luo(罗庆), Tian-Cheng Gong(龚天成), Hang-Bing Lv(吕杭炳)†, Qi Liu(刘琦), and Ming Liu(刘明)

Key Laboratory of Microelectronics Devices and Integrated Technology, Institute of Microelectronics of the Chinese Academy of Sciences, Beijing 100029, *China*

(Received 26 May 2020; revised manuscript received 11 January 2021; accepted manuscript online 28 January 2021)

The resistive random access memory (RRAM) has stimulated a variety of promising applications including programmable analog circuit, massive data storage, neuromorphic computing, etc. These new emerging applications have huge demands on high integration density and low power consumption. The cross-point configuration or passive array, which offers the smallest footprint of cell size and feasible capability of multi-layer stacking, has received broad attention from the research community. In such array, correct operation of reading and writing on a cell relies on effective elimination of the sneaking current coming from the neighboring cells. This target requires nonlinear *I*–*V* characteristics of the memory cell, which can be realized by either adding separate selector or developing implicit build-in nonlinear cells. The performance of a passive array largely depends on the cell nonlinearity, reliability, on/off ratio, line resistance, thermal coupling, etc. This article provides a comprehensive review on the progress achieved concerning 3D RRAM integration. First, the authors start with a brief overview of the associative problems in passive array and the category of 3D architectures. Next, the state of the arts on the development of various selector devices and self-selective cells are presented. Key parameters that influence the device nonlinearity and current density are outlined according to the corresponding working principles. Then, the reliability issues in 3D array are summarized in terms of uniformity, endurance, retention, and disturbance. Subsequently, scaling issue and thermal crosstalk in 3D memory array are thoroughly discussed, and applications of 3D RRAM beyond storage, such as neuromorphic computing and CMOL circuit are discussed later. Summary and outlooks are given in the final.

Keywords: resistive switching memory (RRAM), three-dimensional (3D) integration, reliability, computing

PACS: 85.25.Hv, 87.19.lv, 73.40.Rw **DOI:** 10.1088/1674-1056/abe0c4

1. Introduction

Memory is one of the basic elements in semiconductor integrated circuit and has an increasing demand on density, power consumption, and performance. The current mainstream in nonvolatile memory market is the flash memory, whose density is predicted to be multiplied in every two or three years by Moore's law.[1] The conventional approach for increasing the density is to shrink the pitch size by advanced manufacturing process. However, as the cell size proceeds down to 10 nm, the planer two-dimensional (2D) structure is facing serious challenges in terms of physical limitation of device structure, coupling of line-to-line, etc.[2,3] Besides, the benefit brought by the increased density becomes insufficient to compensate the growing cost of advanced fabrication process. In view of this condition, both academic and industrial communities turn their sight to the vertical integration of memory array with so called BICS (bit cost scaling), TCAT (terabit-cell-array transistor), or VSAT (vertical stacked array transistor) structures.[3–5] The vertical three-dimensional (3D) integration is becoming a key approach to realize high density with tera-bit per cm^2. Among the various new concept emerging non-volatile memories, the resistive switching random access memory (RRAM) owing to its simple structure, good compatibility with complementary metal–oxide–semiconductor transistor (CMOS) process, ease of stacking with the back end of line (BEOL), has received enormous attention from the research community.[6–13] According to the international technology roadmap of semiconductor (ITRS) 2013,[14] the RRAM was thought as a new technology being worthy of additional focus to accelerate the commercialization process.

The RRAM array architecture can be divided into two categories, passive array or active array. In passive array, the memory cells are located at the cross-point of perpendicularly placed parallel wordlines (WLs) and bitlines (BLs). In 2D planer structure, it can achieve the smallest cell size of $4F^2$ (F is the feature size). Because of the independence of the processes of the front-end-of-line (FEOL), the passive crossbar array can be readily stacked along with the process of the BEOL. Given an N layered crossbar array, the cell size per bit will be as small as $4F^2/N$. This highly compact inte-

*Project supported in part by the National Key R&D Program of China (Grant Nos. 2018YFB0407501 and 2016YFA0201800), the National Natural Science Foundation of China (Grant Nos. 61804173, 61922083, 61804167, 61904200, and 61821091), and the fourth China Association for Science and Technology Youth Talent Support Project (Grant No. 2019QNRC001).

†Corresponding author. E-mail: lvhangbing@ime.ac.cn

© 2021 Chinese Physical Society and IOP Publishing Ltd

gration manner owns a promising prospect for high density data storage. In contrast, the active array, with a 1T1R (one transistor one resistor) structure as its general form, can perfectly insulate the accessed cell from neighboring cells by a transistor, which also provides compliance to avoid over-programming. Because of its logic based integration manner, the 1T1R structure becomes a rational choice for embedded application.[15–20] However, the transistor consumes silicon area and is incapable of multi-layer stack, making the 1T1R structure difficult for 3D integration. There are also some reports combining the passive and active integrations together by adopting such a structure named 1TxR with a purpose of better performance and higher density.[21,22] In 1TxR structure, each transistor drives a vertically stacked ninny crossbar array. This architecture can balance the strict requirement of high nonlinearity in passive array and difficulty of 3D integration in active array. Considering the design of 1TxR structure overlapped with the passive array in terms of nonlinearity and leakage current, this structure will not be discussed extensively in this paper.

The major task of designing a 3D passive array is to achieve a low cost, highly compact memory array and to operate each cell in array correctly without disturbing the neighbors. The fabrication cost mostly depends on the array architecture, e.g., stacked array or vertical array. The later one is also called vertical RRAM (VRRAM), with less lithography steps for patterning and thus lower cost. There are various challenges in 3D passive array designing, such as the sneaking current, disturbance, IR drop, thermal coupling, etc, which greatly retard the progress towards commercialization. Apart from data storage, the passive network also has other potential applications, such as neuromorphic computing,[23] 3D CMOS/molecular circuits (CMOL),[24] etc., with promising prospect of low cost and high energy efficiency. In this review article, we will firstly address the associative problems in passive array and 3D architectures (Section 2), and then review the states of the art of various selector devices and self-selective cells (Section 3). Key parameters that influence the device nonlinearity and current density are discussed based on the corresponding working principles. Section 4 addresses the reliability issues in 3D array such as uniformity, endurance, retention, disturbance, etc., and Section 5 discusses the issues of scaling and thermal crosstalk in 3D memory array. Applications of 3D RRAM beyond storage, such neuromorphic computing and CMOL circuit are discussed in Section 6. Summary, discussions, and outlooks are given in the final.

2. Associative problems in passive array and 3D architectures

Passive storage arrays can achieve the highest storage density at a very low cost. However, the leakage current problem in passive arrays could lead to problems of reading errors and disturbance of resistance state in partly selected cells. In addition, the IR drop and thermal coupling become more and more serous as the array size increases.

2.1. Sneaking current problem

The sneaking current path is a fundamental problem commonly existing in 2D or 3D cross-bar array,[25–30] as shown in Fig. 1. In the V scheme, the voltage with magnitude V is applied only to the selected word line, while all the unselected word and bit lines are floating. If the R2, R3, and R4 are at low resistance state, the readout state of the selected R1 will always be low resistance state (LRS) whatever the state of R1 is. The sneak current path consists of two forward biased half-selected cells and a reverse-biased cell. Here, the forward and reverse bias directions are defined as being identical and opposite to the bias at the selected cell. Hence, there are two ideas to solve the leakage problem. Firstly, connecting a diode selector can impede the sneak currents by limiting the current flow through the reverse-biased cell R2. In such a way, the sneaking current from the undesignated cell could be suppressed by adopting a bias scheme of $-V$ method, i.e., the selected cell is biased by V and the unselected cells is bias by $-V$ or 0. However, the diode selector has some inherent limitations, such as area dependence of driving current and inconsistency with the bipolar device.[31–33] Secondly, decreasing the current of the forward-biased cells of R3 and R4 at the low-voltage region by constructing a threshold switch in LRS. Given this, the selector device with symmetrical (or nearly symmetrical) highly nonlinear I–V relationship was put forwards.[34] The nonlinearity (η), defined as the ratio of $I@V_{\rm read}/I@(V_{\rm read}/2)$ is helpful to diminish the sneak current from undesignated cells by using $V/2$ bias scheme. Besides introducing separate selector device, developing self-selective cell (SSC), such as build-in nonlinear device, complimentary switching device, etc, is also a feasible way to solve the sneaking current issue. As will be discussed later, the SSC device has special meaning for the VRRAM designing.

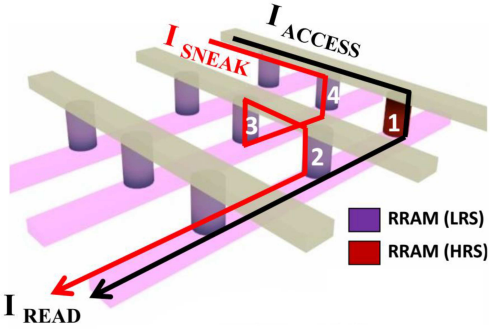


Fig. 1. A schematic diagram of the typical cross-bar showing the read disturbance problem by the presence of sneak current.

2.2. Disturbance issue

The disturbance mainly occurs in the devices which locate in partly selected rows or columns, depending on different bias schemes. In a passive array, the $V/2$ bias scheme is widely adopted in order to reduce the interference of unselected cells. In this scheme, the half-selected cells suffer from consecutive sub-switching voltage and bear the risk of the change of resistance states. This is usually called programming disturb. The disturbance issues are much more complicated in an array. In the cross-point RRAM arrays, because of waveform degradation caused by the parasitic resistance and capacitance of the bit-lines and word-lines, voltage signals larger than the critical switching voltage are usually applied in order to ensure the correct programming for the farthest cell. In this case, half-selected cells nearer to the voltage supply will suffer from larger voltage stresses than $V/2$ and serious disturb issues during the programming operation. The adoption of nonlinear selector could moderate the disturbance issue greatly, because most of the voltage will be dropped on the selector device at low voltage region.

2.3. IR drop issue

The IR drop issue is related with the voltage drop on the WLs or BLs. According to the reported data, the resistance of a metal line is exponentially dependent on its size in nm scale.[38] This makes the resistance of metal line ineligible as compared with the resistance of memory cell. A large potential difference is applied across selected cells by driving the corresponding wires to high voltages. Some part of the applied voltage could be dropped on the metal lines and the unselected cells (Fig. 2), making the voltage applied on the cell lower than expected. The total voltage drop across any selected nonvolatile memory (NVM) device must be sure to switch the device, despite the additional voltage drop in its own access device and in the wiring. The worst case happens in the device located at the farthest corner.

The IR drop issue will not only deteriorate the read/write disturbance, but also increase the sneak current, in the case of the leakage of the selector highly dependent on voltage bias.[39] Increasing the device nonlinearity could be helpful to improve the capability of voltage delivery. Comparison between the voltage distributions along the WL in the arrays with linear and nonlinear devices shows that higher nonlinearity could result in less degradation of delivering voltage.[40] Together with the nonlinearity and on/off ratio, the metal line resistance becomes the main factors on determining the maximum reachable array size.

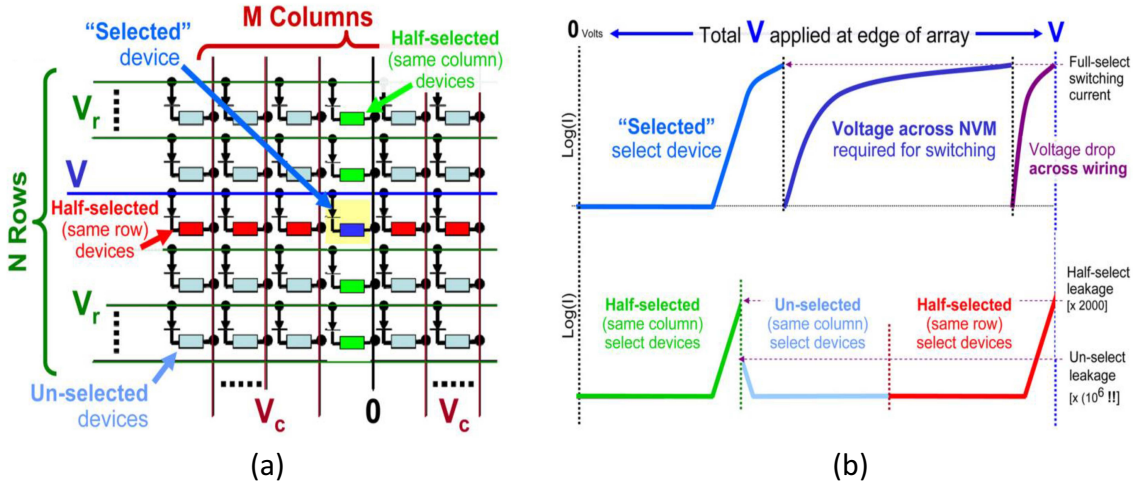


Fig. 2. (a) In a cross-point array of NVM devices, three other sets of NVM/access devices would be affected when selecting one NVM/access device pairs, including the half-selected cells and unselected cells, as well as the wires. (b) The total voltage drop across any selected NVM device must be sure to switch the device, despite the additional voltage drop in its own access device and in the wiring. However, that same applied voltage must not lead to excessive leakage in either the half-selected or unselected devices. Reproduced with permission from Ref. [40].

2.4. Thermal coupling

The thermal coupling is related with the resistance state changes in the neighbor cells resulting from the high temperature generated during programming. Despite the resistive switching mechanism is related with several aspects, there is a common sense that Joule heating or thermal effect plays an important role in it, regardless the switching type.[43–47] Temperature as high as hundreds of degrees could be generated inside the device during programming. The large amount of Joule heat will cause the resistance degradation in adjacent memory cells of the RRAM crossbar array. On one hand, the thermal dissipation of a programmed cell depends on the thermal conductivity of its surrounding material, such as insulating layer and metal line. One the other hand, it is also influenced by the structural factors, such as 3D architecture and selector device. Higher thermal conductivities of the insulating material, a larger LRS resistance, and a faster operational speed were found to effectively reduce the thermal disturbances in 3D one

selector one resistor (1S1R) RRAM crossbar arrays. Since it is difficult to directly monitor the temperature profile produced inside the array during programming, the development of analytical model is helpful to analyze the thermal issue and figure out a guideline for structure designing.

2.5. The 3D architecture

The 3D RRAM is considered as one of the strongest contenders for NAND due to its aggressive scalability. Two typical 3D architectures, the 3D vertical RRAM (VRRAM) architecture and 3D horizontal RRAM (HRRAM) architecture, have been proposed. The former one is deriving from 3D vertical NAND architecture and process, while the latter is realized by stacking planar cross-point arrays layer by layer.

The HRRAM is believed to have higher lithograph cost since the numbers of lithography masks and process steps are dependent on the number of stack layers. In comparison, the VRRAM is the most economical as the channel hole is one-time implemented. However, the 3D VRRAM is not expected to scale aggressively with the advanced technology since the reliability is relatively dependent on the diameter of the channel hole. Besides, it puts forwards higher requirements on material etching and deposition process, adding the cost of fabrication. The HRRAM is considered free from the scaling limitation due to the simple fabrication process. Single cell lateral scalability below 10 nm and vertical scalability below 5 nm have already been demonstrated, thus very small cell area and ultra-high memory density can be achieved in 3D HRRAM.

3. Selector devices and self-selective cells

3.1. Nonlinear selector devices

An ideal selector should have several requirements for application in stacked cross bar array. In each selected cell, the selector should be capable of supplying high current to meet the needs of programming and erasing the memory element in series, typically on the order of several MA/cm.[2,29,36,39,40] The aggregate leakage through all the unselected cells should be as low as possible, since the number of unselected cells is much larger than that of the selected cells, thus dominating the power consumption of the entire array. The first two points mentioned above (high on-state current and low off-state leakage) means that the selectors should have a highly nonlinearity. To satisfy the needs of the RRAM device which is operated in bipolar mode, the selector should also be capable of bidirectional operation. Furthermore, since the selector is used for stacked cross bar array, it should be fabricated by BEOL compatible process. The property of the selector should not be a limitation of RRAM performance, including switching speed, endurance, variability, and so on. All the requirements listed above are the targets of the selector. Various selector devices will be discussed in following in terms of their strengths and weaknesses.

Selectors with different structures and different theories have been reported in the past few years. Two-terminal NPN diodes have been proposed as selector.[53,58–61] Srinivasan *et al.* obtained a maximum ON/OFF ratio ($> 10^6$) and high ON-state current density (> 1 MA/cm^2),[53] as shown in Fig. 3(a). However, the epitaxial growth process and the crystalline Si template require high temperature (~ 700 °C) process. Such a device is not BEOL-compatible and incapable of stacking above the Si substrate. This condition has motivated research into non-Si-based selector with a low temperature process. Non-Si-based selectors can be divided into several groups according to their mechanisms, such as oxide diodes, threshold switch, oxide tunnel barrier, and mixed-ionic-electronic-conduction. Several groups have studied diodes made from semiconducting oxide hetero-junctions[48,62,63] and two-terminal metal–insulator–metal (MIM) structures.[54,64] In these works, the most considerable one was reported by Huang *et al.*, with Ni/TiO_2/Ni structure, having an ON/OFF ratio of ~ 4700,[54] as show in Fig. 3(b). A small 8×8 crosspoint 1S-1R array was demonstrated by implementing Ni/TiO_2/Ni selector with bipolar RRAM. This selector device has advantages of high temperature (125 °C) operation and being capable of scaling, however, the endurance and disturbance were not mentioned.

The threshold switching devices with high selectivity and low off-current are essential to suppress the sneak current from the neighboring cells which can lead to failure of write and read operations. It turns on from an initial low conducting state to a highly conducting state as soon as a threshold voltage or current is applied. Various types of threshold switches have been proposed in literatures.[55,56,65] As shown in Fig. 3(c), in a nanoscale VO_2 device, the current through it increases suddenly at 0.35 V as the conductivity of VO_2 changes from an insulating state to a metallic one.[55] If the voltage is reduced below 0.2 V, the current decreases suddenly as the VO_2 returns to its initial insulating state. Furthermore, this device shows very fast response (< 20 ns), stable and uniform switching and holding voltages. However, for practical application at high temperatures up to 85 °C, it fails because of the low transition temperature (~ 67 °C). Another metal insulator transition (MIT) selector (Pt/NbO_2/Pt) with higher transition temperature shows excellent thermal stability up to 160 °C, and exhibits well-behaved and repeatable hysteretic *I*–*V* characteristics over 1000 switching cycles (Fig. 3(d)).[56] However, the leakage of this structure was high and only limited nonlinearity was achieved. Figure 3(e) shows the typical *I*–*V* curve of a threshold vacuum switch (TVS). This device exhibits high current density (> 100 MA/cm^2), high selectivity of $> 10^5$, and good endurance of over 10^8 cycles using 100 ns programming

pulses. However, array level demonstration was not reported and the variation was an issue of big deal considering that the tunneling current is highly dependent on the ultranarrow vacuum gaps. Oxide tunnel barrier is an attractive approach for achieving strongly nonlinear *I*–*V* characteristics. Tunnel barriers can be formed by high-*k* materials (HfO_2, ZrO_2, and TiO_2).[34,57] Luo *et al.*[34] reported TaO_x-based selector with trapezoidal band structure ($Ru/TaO_x/TaO_y/W$), as shown in Fig. 3(f). Low off-state current ($\sim$ 10 pA) and large nonlinearity ($\sim 5 \times 10^4$) have been achieved. Excellent uniformity was achieved during 10^3 successive DC cycles. No disturbance was observed under read and half-read conditions during 10^{10} pulse endurance test. This selector was fabricated in 1 kb crossbar array and the integrated 1S1R device showed high nonlinearity in low resistance state (LRS).

Copper-containing mixed ionic-electronic conduction (MIEC) materials have been proposed for 3D-ready selector.[66–69] As demonstrated in the work from IBM research group, MIEC-based selectors offer high nonlinearity (10^7), ultralow leakage ($<$ 10 pA), high voltage margin V_m, and excellent turn-on speed ($<$ 50) (Fig. 4(a)). MIEC materials contain a lot of mobile copper ions, which can move readily under an electrical field. At low bias, the Schottky barrier at the interfaces between MIEC and electrode, strongly suppressing the hole conduction, leads to the low leakage. As the bias increases, copper ions move toward the negative electrode, modulating the interfaces and leading to an exponentially increase in electronic current. Figure 4(b) shows the endurance of the MIEC-based device. As the compliance current increased, the endurance of this device was degraded. Good endurance as high as 10^6 was achieved in the case of high operating current ($\sim$ 150 μA). Excellent uniformity was observed according to the cumulative distribution functions (CDFs) of the *I*–*V* characteristics in an array (Fig. 4(c)). MIEC based access device could maintain low-leakage over hours of exposure, both in a deep ($\pm$230 mV) and shallow un-select ($\pm$350 mV) condition (Fig. 4(d)). To demonstrate access characteristics of MIEC devices, a PCM+MIEC device was formed. Implementation of MIEC selector into phase change memory crossbar array in a form of 1S1R structure was successfully demonstrated. As shown in Fig. 4(e), more than 10^6 switching cycles with RESET currents $>$ 200 μA and 5 μs-long SET pulses ($\sim$ 90 μA), were achieved, indicating that the MIEC device has a hopeful prospect for selector use. Further improvements would include a higher endurance at high current and increase the voltage margin so as to enable large arrays for NVM requiring more than 1.5 V switching voltage.

Field-assisted super linear threshold (FAST) selector was reported by Jo *et al.*[70] Excellent selector performance was presented such as selectivity of 10^{10}, sub-50 ns operation, and 100M endurance. Figure 5(a) is the typical *I*–*V* curves of a FAST selector. The threshold voltage (V_{th}) of the selector can be turned by controlling the thickness of the active layer. The turn-on speed is less than 50 ns under applied voltage above V_{th} and the on-to-off transition time is less than 5 ns (Fig. 5(b)). The selector can recover quickly to the off-state when the voltage is removed, with a recovery time $<$ 50 ns (Fig. 5(c)).

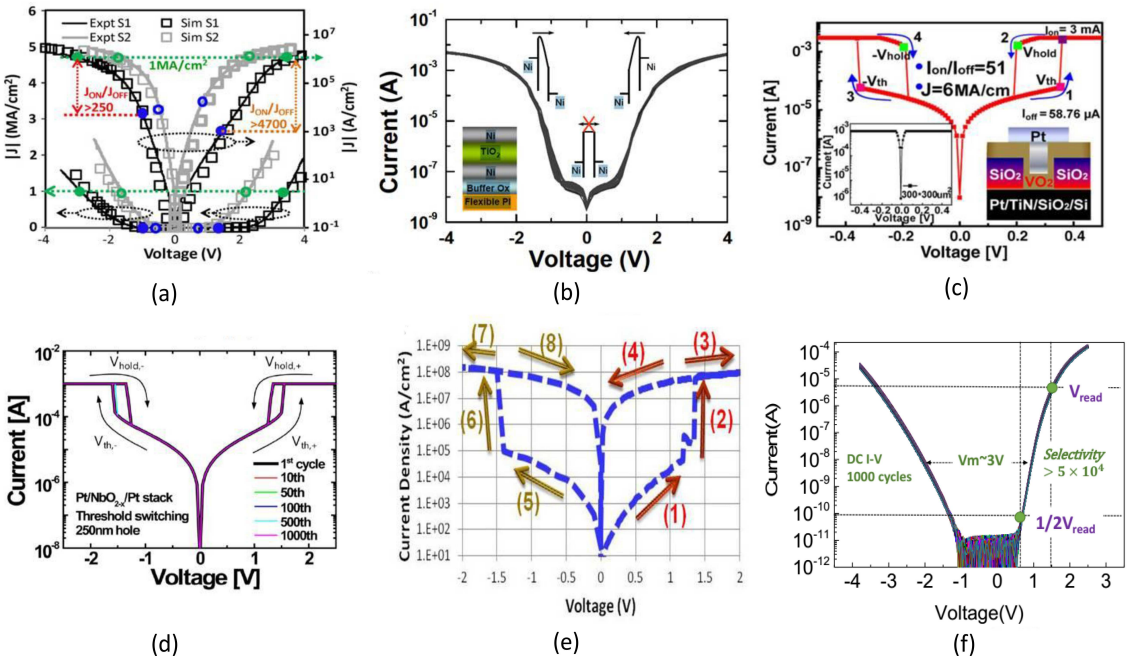


Fig. 3. (a) Current density vs. voltage plots for two NPN structures with different doping profiles. (b) Bidirectional exponential current–voltage curves measured on the devices with the structure of $Ni/TiO_2/Ni$. (c) *I*–*V* characteristics of threshold selector of the $Pt/VO_2/Pt$ device. (d) *I*–*V* characteristics of nanoscale niobium dioxide device structured as $Pt/NbO_2/Pt$. (e) *I*–*V* characteristics of TVS selector. Arrows (1)→(8) represent the measurement sequence. TVS volatile switching behavior provides high current density and selectivity. (f) Typical *I*–*V* switching characteristics of a tunnel barriers selector ($Pt/TaO_x/TiO_2/TaO_x/Pt$). Reproduced with permission from Refs. [34,53–57].

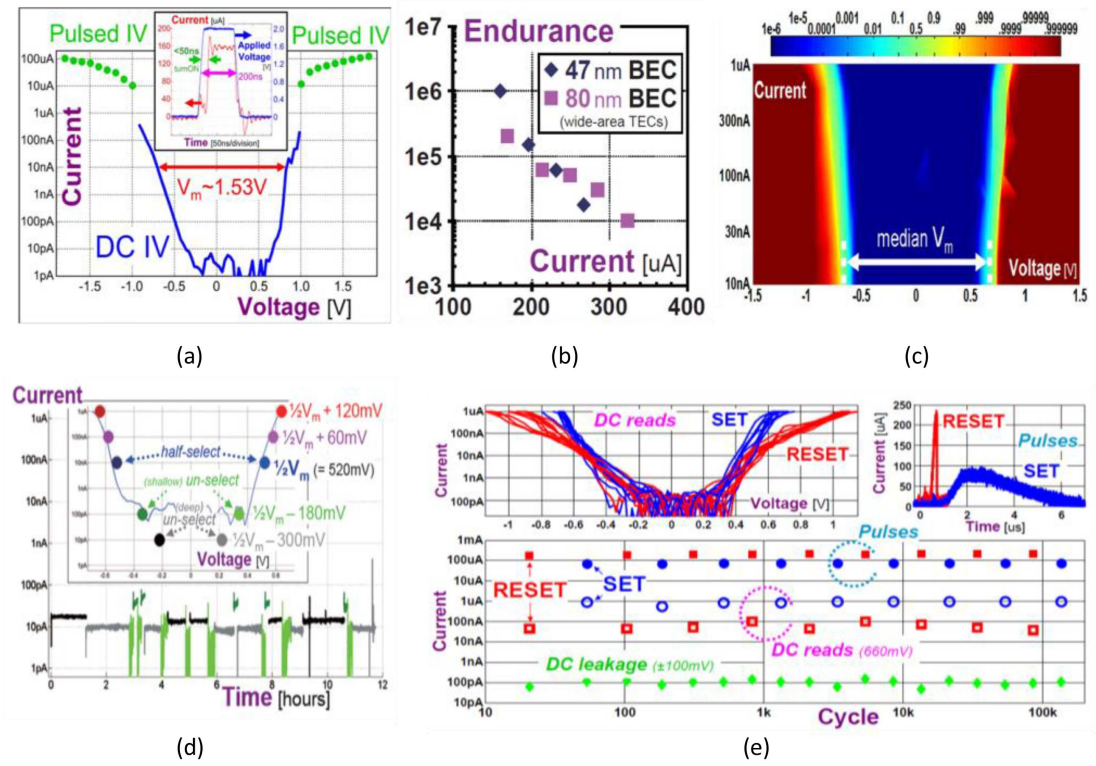


Fig. 4. (a) Typical *I*–*V* curves of MIEC selector showing $> 10^7$ nonlinearity and < 50 ns turn-on switching (insert). (b) MIEC-based selector endurance depends on current, but is independent of bottom electrode contact CD. (c) The *I*–*V* results with tight distributions across a 512 × 1024 array of integrated MIEC access devices. (d) MIEC selectors exhibit highly immunity in a deep (±230 mV) or shallow un-select (±350 mV) condition. (e) The PCM and MIEC device-pair shows excellent endurance up to 100k cycles. Reproduced with permission from Refs. [66–69].

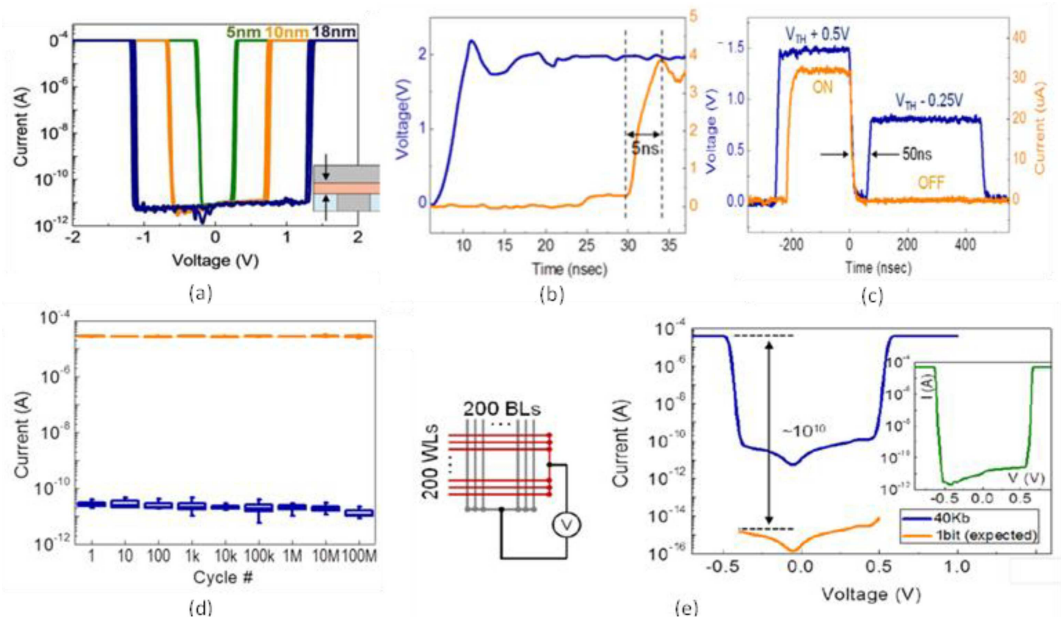


Fig. 5. (a) The *I*–*V* characteristics of FAST selectors with different thicknesses of the active layer. (b) Off-to-on transition time is about 5 ns (tester limited) for the selector. (c) The selector quickly relaxes to the off state in less than 50 ns. (d) The selectors can be reliably cycled over 100M cycles while maintaining $> 10^6$ on/off ratio (test limited). (e) The 1-bit (100 nm × 100 nm) leakage current can be projected through entire 40 kb devices. Inset: Typical *I*–*V* of a selector on the same wafer. Reproduced with permission from Ref. [70].

Figure 5(d) shows that the device can reliably switch over 100M cycles. To extract the intrinsic leakage current of the FAST selector, the leakage current through an entire 40 kb selector array was measured (Fig. 5(e)). The extracted selectivity is extrapolated to be as high as 10^{10} (@100 nm device). The FAST selector almost meets all the requirements that an ideal selector should have. However, the intrinsic mechanism and the compatibility with CMOS process have not been re-

vealed yet. The similar threshold switching selector based on Cu:HfO_2 was proposed by Luo,[71] which shows high nonlinearity and ultra-low off-state. The threshold switching behavior is originated from the spontaneously rupture of the unstable conductive filament. Zhao *et al.*[72] proposed a high driving current selector by utilizing graphene defect engineering in the structure of Pt/SiO_2/graphene/Ag. This selector exhibited excellent bidirectional selectivity as high as 5×10^8 under high complicate current and demonstrated the ability to suppress the crosstalk issue in large memory arrays.

3.2. Self-selective devices

The 3D vertical RRAM provides a cost-effective approach for high density integration. However, unlike 3D stacked crossbar array, individual selector is not allowed in VRRAM. Developing high performance self-selective cell (SSC) is the most critical issue for the integration of vertical RRAM. Some attractive device configurations have been proposed to realize self-selective cell, such as build-in nonlinearity device, self- rectifying device, and complementary resistive switching device.

Park *et al.* reported a resistive switching memory with build-in nonlinearity by inserting a thin barrier layer between oxide and electrodes.[50,75] The device exhibited a nonlinearity about 17 under low compliance current of 1 μA (Fig. 6(a)) and could be repetitively switched between LRS and HRS for more than 10^7 cycles with 1 μs voltage pulses, maintaining an on/off ratio about 100. Figure 6(b) shows the non-linear *I*–*V* characteristics of the Ti/HfO_x/TiO_x/Pt RRAM device with higher compliance current ~ 50 μA and 20 μA. Lightly increased nonlinearity about 23 was observed. However, it is still not large enough to prevent the leakage of the unselected cell in a large crossbar array. Son *et al.* reported a hybrid memory device which combined metal–insulator transition and memory device together by controlling oxygen concentration in vanadium oxide film.[73] Vanadium oxide film with various intermediate oxygen content showed hybrid characteristics, with nonstoichiometric V_2O_{5-x} film showing typical resistive memory switching behavior and stoichiometric VO_2 film exhibiting typical MIT behavior. As shown in Fig. 6(c), this device exhibited both threshold switching and bipolar memory switching behavior and had nonlinearity about 80. Similar to vanadium oxide but offering a higher transition temperature, niobium oxide exhibits both memory and selection properties for intermediate oxygen concentrations.[74] Figure 6(d) shows the typical *I*–*V* curve of W/NbO_x/Pt memory device. The LRS of the memory device offers a nonlinearity of 10^2. Without additional selector device, 1 kb cross-point array based on the hybrid memory device with successful SET/RESET operation up to 10^6 cycles was demonstrated. However, all the devices mentioned above were not integrated on 3D V-RRAM and further improvements are still needed for application of large-scale array.

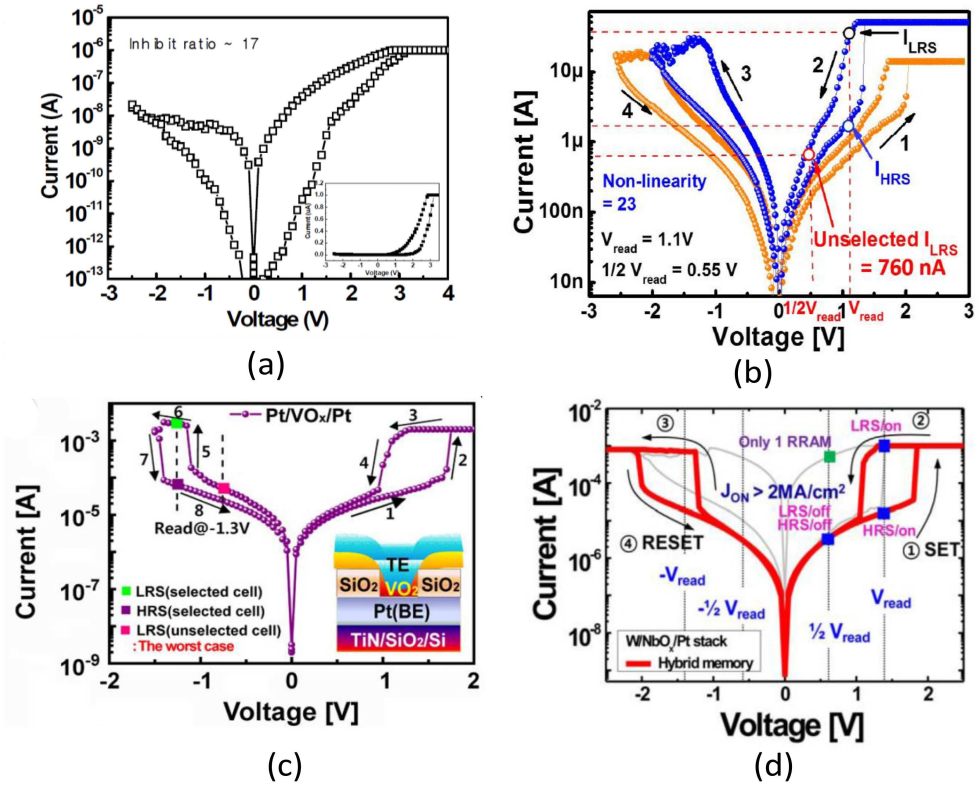


Fig. 6. (a) The *I*–*V* curve of the structure of TE/TMO/BE. (b) Typical *I*–*V* curves of the self-selective RRAM (Ti/HfO_x/TiO_x/Pt). (c) *I*–*V* curve of the hybrid-type Pt/VO_x/Pt device. Inset: Structure of the nanoscale device. (d) *I*–*V* characteristics of hybrid (W/NbO_x/Pt) memory device. Reproduced with permission from Refs. [73–75].

Luo *et al.*[76] successfully demonstrated a self-selective device with extremely high nonlinearity of 1000 and ultra-low leakage current (0.1 pA). The hybrid devices combine HfO_2 layer as the switching layer and mixed ionic and electron conductor (MIEC) layer as the selective layer. Based on this SNR device, a four-layer 8×32 VRRAM 3D array was successfully demonstrated. Figure 7(a) is the schematic of 4-layer VRRAM and SEM images of in-line checks. After multiple TiN(60 nm)/SiO_2(100 nm) layers deposited respectively, patterning and one-step etching were applied to form stacked wordlines (WL). HfO_2/CuGeS bilayer and top electrode were deposited on the sidewall, followed by liftoff process to form bitlines (BL). Figure 7(b) is the TEM image and EDX line scan of the device. Well-distinguishable layers with clearly identified interface could be observed. Figure 7(c) shows the typical *I*–*V* characteristics of HfO_2/CuGeS bilayer device. The half-select leakage as low as 0.1 pA and nonlinearity as high as 10^3 were achieved. In addition, the operation current below nA is beneficial to realize low power consumption and high-density storage. As shown in the insert of Fig. 7(c), the programming voltage of the SSC can be modulated by the thickness of the CuGeS layer. Figure 7(d) is the typical *I*–*V* sweeping curves of the cells in 4-layer VRRAM array. Each layer device exhibits stable and uniform self-selective memory characteristics. Other excellent characteristics such as high endurance ($> 10^7$) and robust read/write disturbance immunity were also presented.

The self-rectifying characteristics within the individual cell is another approach to suppress the sneak current in high density array in 3D VRRAM. The *I*–*V* characteristics of the self-rectifying is similar as the one diode one resistor (1D1R) structure, showing the obvious rectification in low-resistance state. The rectification is originated from the Schottky barrier created neither at the interface of electrode and switching layer or at the conjunction of filament and buffer layer. Initially, semiconducting electrodes such as n-Si or n-Ge were usually used,[73,77] with corresponding *I*–*V* curves shown in Fig. 8(a). Although the rectifying ratio for these devices was reported as high as 10^6, the Si or Ge based self-rectifying device has inherent drawbacks, due to the high processing temperature, making them incapable to be stacked layer by layer in BEOL process.[78] In order to achieve a stackable self-rectifying device, semiconducting buffer layer was introduced between the metal electrode and switching layer, e.g, a-Si/WO_x/Pt structure proposed by Lv *et al.*[79] In this device, the semiconducting material (a-Si) acts as the role of c-Si or c-Ge, while having lower processing temperature and stackable capability (Fig. 8(c)). However, the rectifying ratio was not satisfactory because of the large amount of interface traps (Fig. 8(d)). Yoon[80] proposed a two-layered device, where a Pt/Ta_2O_5/HfO_{2-x}/TiN with high uniform, forming free, and self-rectifying characteristics was proposed (Fig. 8(e)).

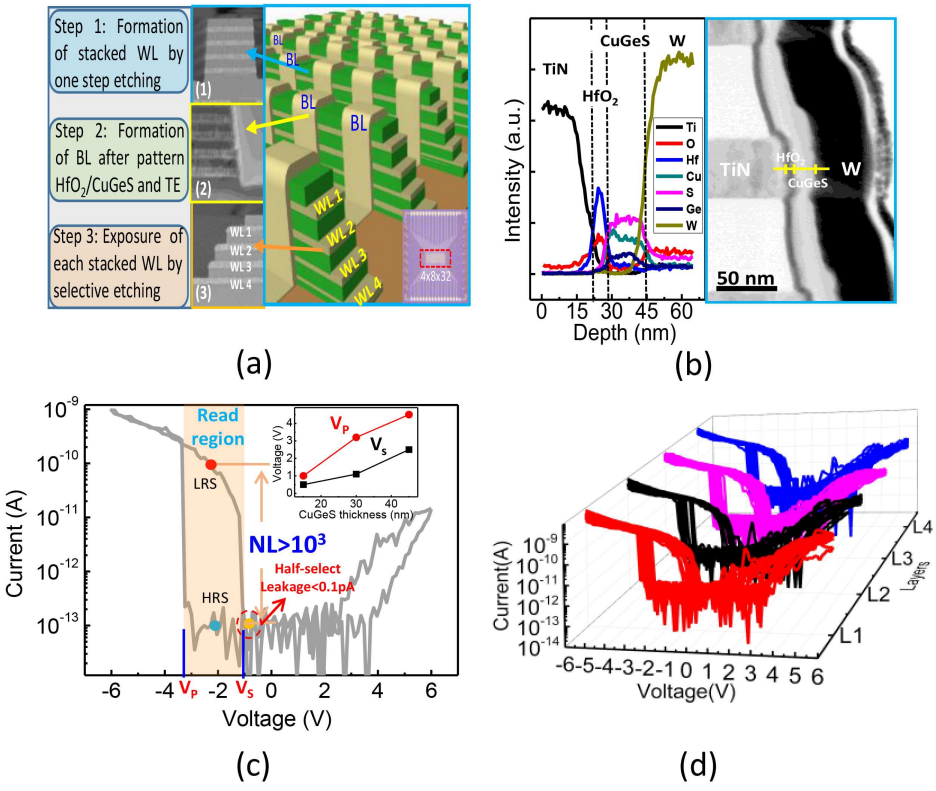


Fig. 7. (a) The process integration of 4-layer 3D V-RRAM. (b) TEM image and EDX line scan over the cross-section of the SSC device (TiN/HfO_2/CuGeS/W). (c) Typical *I*–*V* curve of the bilayer SSC. Insert: The dependence of the programming voltage (V_P) and selective voltage (V_S) on the thickness of CuGeS. (d) The devices exhibit stable and uniform *I*–*V* curves in the 4-layer 8×32 arrays. Reproduced with permission from Ref. [76].

The working principle of this device is that the HfO_x layer works as resistive switching layer by electron trapping and de-trapping process, while the other dielectric layer Ta_2O_5 remains intact during the switching cycles and creates a high Schottky barrier with the metal electrode (such as Pt) of high work function. The high uniformity takes the advantage of electronic switching mechanism, by avoiding uncontrolled ions migration. The most recent progress achieved on self-rectifying device came from Li's work,[81] a fully foundry-compatible, all-silicon-based and self-rectifying memristor with a p-Si/SiO_2/n-Si structure was demonstrated. The device exhibits 10^5 rectifying ratio, 10^4 on/off ratio, and excellent retention at 300 °C. The 3D crossbar arrays up to five layers were further build.

The complementary resistive switching (CRS) device consists of two back-to-back-connected bipolar resistive switches. The concept of CRS was firstly proposed by Waser and co-workers,[80–82] its schematic and typical I–V curve are depicted in Figs. 9(a) and 9(b), respectively, based on conductive bridging memories (CBRAM) as the switching elements. The anti-series connected pair exhibited four possible combinations of HRS and LRS, as shown in Fig. 9(c). Beginning with an original HRS/HRS state, the HRS/LRS can be defined as the state '0' and LRS/HRS as the state '1'. The read operation was performed under a read voltage between V_{th1} and V_{th2}. The leakage at half read voltage was determined by HRS, thus greatly suppressing the sneak current. As the stored information was '0', the read voltage had no influence on both top and bottom cells. However, as the stored information was '1', the read operation of the CRS cell was destructive, because the read voltage higher than V_{th1} could trigger the bottom cell into

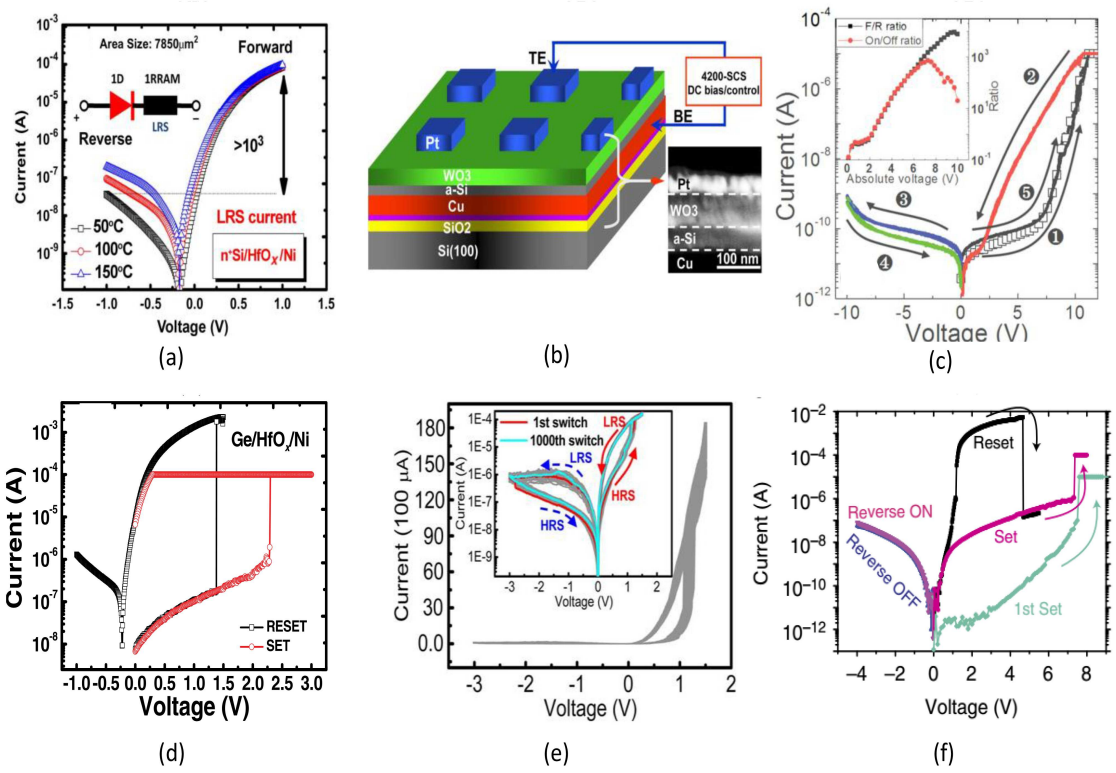
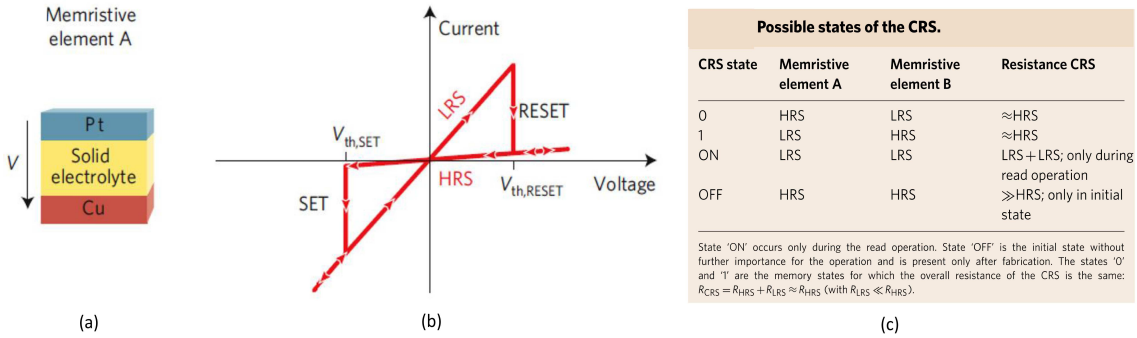


Fig. 8. (a) The I–V curves of n-Si/HfO_x/Ni device. (b) Schematic of Cu/a-Si/WO_3/Pt device. (c) I–V curves of Cu/a-Si/WO_3/Pt device. (d) I–V curves of Ge/HfO_x/Ni device. (e) I–V curves of TaO_x/HfO_2 bilayer self-rectifying device. Insert: The rectifying and On/Off ratio at different applied voltages. (f) Electrical behavior of p-Si/SiO_2/n-Si self-rectifying device. The ON/OFF ratio was higher than 10^4 and rectifying ratio $\sim 10^5$ when bias voltage was larger than −1.5 V. Reproduced with permission from Refs. [77,79,81,82].



Possible states of the CRS.

CRS state	Memristive element A	Memristive element B	Resistance CRS
0	HRS	LRS	≈HRS
1	LRS	HRS	≈HRS
ON	LRS	LRS	LRS + LRS; only during read operation
OFF	HRS	HRS	≫HRS; only in initial state

State 'ON' occurs only during the read operation. State 'OFF' is the initial state without further importance for the operation and is present only after fabrication. The states '0' and '1' are the memory states for which the overall resistance of the CRS is the same: $R_{CRS} = R_{HRS} + R_{LRS} \approx R_{HRS}$ (with $R_{LRS} \ll R_{HRS}$).

(c)

Fig. 9. (a) Schematic of a CRS device with back-to-back connected two CBRAM cells. (b) Typical I–V curve of CRS device. (c) Two stored data states amongst the four possible combinations of individual device states. Reprinted with permission from Ref. [80].

LRS. In this case, the read-out current was high (On state), given both top and bottom cells being at LRS. After reading, a write-back process was required to return the CRS device back to state '1'. This partly destructive read out characteristics could be one of the inherent drawbacks of CRS device, which could increase the complexity of peripheral circuitry designing. In contrast to the CRS devices with connecting metal–dielectrics–metal stack in series, the one proposed by Nardi *et al.* only contained single layer structure.[81] By selectively forming the filament at the top or bottom interface, bipolar CRS behavior was successfully demonstrated. Similar phenomenon was also reported in other literatures.[82–86]

4. Reliability issues of RRAM

High-density integration of RRAM cells is the precondition for its future commercialization. Since the yield, density, and operating property of the memory chip are limited by the performance characteristics of the worst cell, the intrinsic uniformity and reliability characteristics of memory devices become critically important. For peripheral circuitry techniques used in the memory hierarchy, a high homogeneity of the devices is critical. The endurance is an important reliability characteristic because higher endurance broadens the application area where frequent read/write operation is required. Retention (stability) is crucial for the nonvolatile memory operation and restricts the intended application of RRAM, such as in code storage, space, high temperature application, and FPGA. The potential scalability to nanometer regime of RRAM devices plays a key role in competitive advantage compared with the traditional Flash memory. Without critical scaling technology, multi-level operation is a desirable capability for high-density memory application. Table 1 summarizes the performance of the RRAM device in 3D architecture reported in recent years. In this section, we will discuss these characteristics of RRAM in terms of 3D architecture.

Table 1. The performance of the RRAM device in 3D architecture reported in recent years.

Structure	3D architecture	Endurance	Retention	On/Off Ratio	Density	Layers	Speed	Reset current	Ref.
Pt/HfO_x/TiON/TiN	vertical cross-point	$> 10^8$	$> 10^5$ s @125 °C	> 10	1 Mb(e)	–	< 50 ns	< 50 μA	[84,87]
Pt/Al_2O_3/Ta_2O_5/TaO_y/Pt	vertical cross-point	$> 10^{10}$	$> 10^5$ s @125 °C	> 1000	10 Mb(c)	–	~ 100 ns	500 μA	[88]
Ta/TaO_y/TiO_2/Ti	vertical cross-point	$> 10^{15}$ @100 °C	$> 10^6$ s	> 10	–	–	–	< 1 μA	[89]
Pt/Ta_2O_5/TaO_y/TiO_2/Pt	1S1R cross-point	$> 10^6$	$> 10^4$ s @85 °C	> 6000	9.5 Mb(c)	–	~ 100 ns	300 μA	[90]
W/NbO_x/Pt	vertical cross-point	$> 10^6$	–	100	32 kb(c)	–	~ 22 ns	2 μA	[91]
TiN/NbO_2/Nb_2O_5/W	vertical cross-point	$> 10^5$	$> 10^4$ s @85 °C	10	1 kb(c)	–	50 ns	2 μA	[63]
Pt/Al_2O_3/HfO_x/TiN	vertical cross-point	$> 10^6$	$> 10^5$ s @120 °C	100	–	–	30 ns	500 μA	[92]
AI/PI-PCBM/Al	stacked cross-point	$> 10^3$	$> 5 \times 10^6$ s	> 1000	8×8(e)	1	100 ms	1 mA	[93]
Ru/TaO_x/W+Cu/HfO_x/Pt	1S1R cross-point	$\sim 10^{10}$	$> 5 \times 10^4$ s	10	> 1 Gb(c)	–	100 ns	100 μA	[50]
TiN/TiO_x/HfO_x/Ru	vertical cross-point	$\sim 10^{10}$	$> 10^4$ s	> 100	100 Mb(c)	4	100 ns	0.1 μA	[51]
Si/SiO_2/Si	stacked cross-point	100	$> 5 \times 10^5$ s @300 °C	$> 10^4$	1 Mb(e)	5	7.6 μs	5 mA	[95]
TiN/HfO_x/TaO_x/Ti/TiN	vertical cross-point	$> 10^7$	> 3 years	> 100	2 Mb(e)	8	100 ns	0.1 μA	[96]

(e): experimental data; (c): calculated data.

4.1. Uniformity

The significant parameter fluctuation is a major barrier to the practical application of RRAM. The fluctuant parameters of RRAM include the operation voltages and the resistances of HRS and LRS, which common exist not only in cycle to cycle, but also in device to device.[87–90] Since the large device-to-device variation of the resistances or switching voltages may cause the read/write failure, the uniformity of the switching parameters is more crucial for the array operation demonstration. The uniformity issue of RRAM is originated from the stochastic oxygen vacancies/ metal ions motion.[8] Chen[91] demonstrated the 3D cross-point architecture with stacking multi-layer vertical RRAM. They showed the consistent switching characteristics among the cells in different layers. The LRS and HRS are about 100 kΩ and 1MΩ, respectively, with the HRS/LRS ratio around 10. The set/reset voltages of the samples in three layers almost coincide with each other, indicating high uniformity of the 3D memory cells.

In filament-type RRAM cells, the oxygen vacancies/metal ions deposition and dissolution are commonly utilized to delimit the resistive switching operation.[8] During the set process, the intrinsic erratic filament (e.g., size/ number and location) causes the fluctuation of R_{LRS} and the reset voltage. Meanwhile, the indeterminate tunneling gap formed in the RESET process is responsible for the variation of R_{HRS} and subsequent SET voltage. It is also found that the LRS/HRS variations are influenced by the oxygen vacancies/ metal ions concentration inside the filament region/tunneling gap region.[92] The unstable switching parameters of RRAM cells must be solved before practical application. Many efforts have been implemented for this issue, such as electrode/oxide interface engineering, location of conductive filament paths, and utilizing optimized operation method.[84,87,93,96–99]

It is well known that embedding appropriate buffer layers in electrode/oxide interface is an effective approach to optimize the RRAM performance. The additional layer could play a role of an oxygen reservoir. By embedding the Al layers, the uniformity of switching parameters of the Pt/AlO_y/HfO_x/TiN

based vertical RRAM devices is remarkably improved compared with Pt/HfO_x/TiN devices.[100] Figures 10(a)–10(c) show the resistance and switching voltage distribution in temporal and spatial scale. Both the cycle-to-cycle and device-to-device uniformity are improved with AlO_y incorporation. The improvement is attributed to that the Al atoms diffused into the HfO_x films are intended to significantly increase the density of oxygen vacancies, thus stabilize the generation of conductive filaments. Utilizing the excellent uniformity, the AlO_y/HfO_x-based devices in the 3-D cross-point array (Figs. 10(d)–10(f)) exhibit stable read and write operations.[101] Each memory cell can be access randomly in the fabricated 3D vertical bilayer array.

Lee *et al.*[102] proposed that the extrinsic defects-related non-uniform switching in HfO_x-based RRAM was significantly minimized by importing the AlO_x layer. Due to the active role of a thin Al_2O_3 buffer layer, highly uniform and reproducible bipolar resistance switching cycles in robust NbAlO-based RRAM could be obtained.[96] Similar optimization was achieved by embedding single-layer graphene,[103] Al layer,[100,104] Ti layer,[99] and CrO_x layer[105] in HfO_x-based RRAM.

CFs paths could also be confined by introducing a certain amount of impurities in the switching layer. Because oxygen vacancy (V_O) can be easily formed near the doping sites where the active energy (E_a) is lower, introducing appropriate dopants (the trivalent metals) can modulate the concentration and location of V_O and achieve fixed CFs along the doping site. Gao *et al.*[106] have successfully improved the dispersion of the forming/SET voltage, LRS/HRS resistance based on the Al-doped HfO_2 RRAM devices, proving the validity of the controllability of trivalent elements doping. Other dopants such as Gd,[89,107,108] Ru,[109] Co,[110] Cr,[97] and Ti[99] are widely used. In other aspects, introducing nanoparticles could significantly improve the uniformity of the RRAM by enhancing the electrode field on the nanoparticles due to the lightening rod effect. Introducing the nanoparticles either in storage media or in electrode material would be practicable for the optimization. Liu[111] proposed that the insertion of Cu nanocrystal (NC) layer effectively improved the switching properties in the ZrO_2-based resistive switching memory. Similar phenomenon was observed in SiO_2-based RRAM by embedding Cu-particles. Kang *et al.*[112] have significantly improved the switching uniformity and device yield of HfO_x-based RRAM devices by decorating the storage layer with Pd nanocrystals. The Ag, Pt, and Ti nanoparticles[5,8,16] are also intentionally selected for optimizing the performance.

Despite the advantages, the material engineering approaches above may increase the fabrication complexity and cost. Optimization methods with respect to operation scheme are also effective in reducing the parameter fluctuations.[87,93,113–115] Liu[114] proposed a novel programming scheme named gate induced resistive switching to improve the uniformity of RRAM in 1T1R structure. The newly proposed scheme ramps the gate voltage while keeps the source and drain voltage constant (Fig. 11(a)). Abrupt RESET was always observed, as can be seen in Fig. 11(b). The possible reason for the abrupt RESET was attributed to the increased trend of power generation by the V_G ramp program scheme, resulting in a positive feedback of Joule heating generation and the accelerated RESET process (Fig. 11(c)). Owing to the effective elimination of the intermediate resistance states by V_G ramp scheme, the uniformity of HRS is greatly improved, as shown in Fig. 11(d).

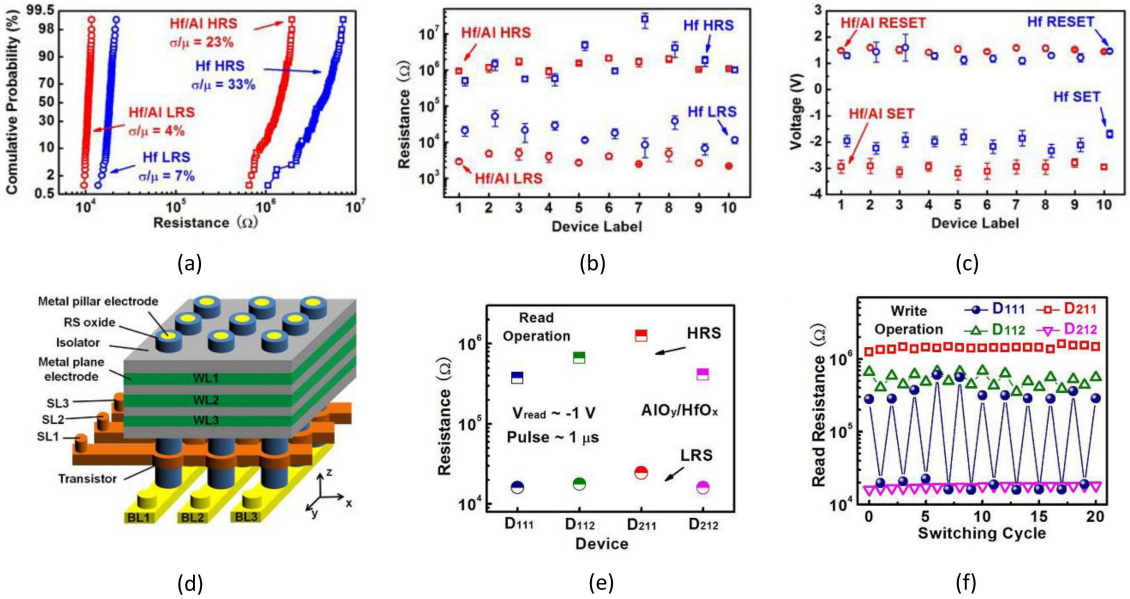


Fig. 10. (a) Comparison of the R_{HRS}/R_{LRS} distribution in 100 consecutive pulse cycles. (b) R_{HRS}/R_{LRS} extracted form 10 different devices. (c) Device-to-device comparison of the SET and RESET voltage distributions. (d) Schematic of the bit-cost 3D cross-point architecture using vertical RRAM array. The vertical MOSFET array as the BL selector addresses the individual cell in the array. (e) The read operation can be executed for the four RRAM devices in the 2×2 vertical array. (f) The set/reset disturb of devices in vertical array. Reprinted with permission from Ref. [101].

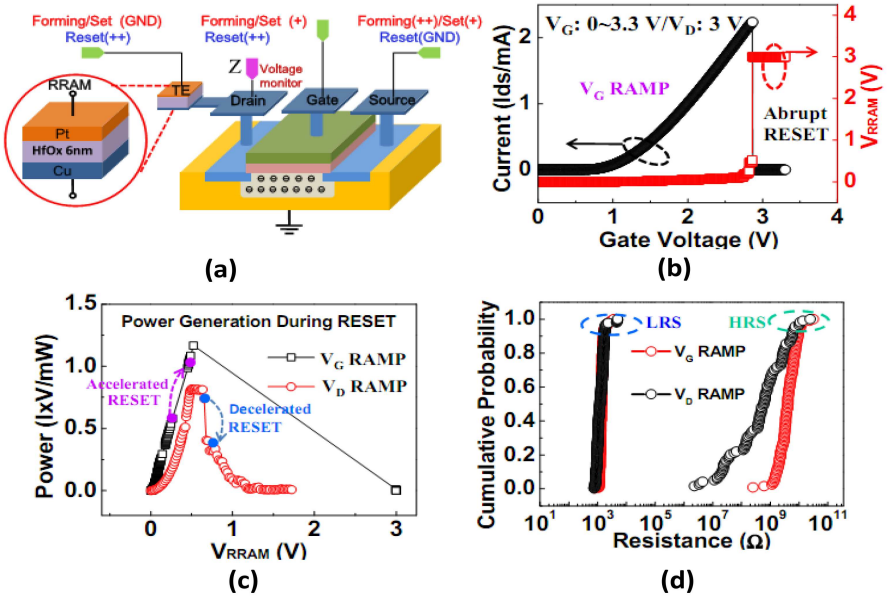


Fig. 11. (a) Schematic of HfO_2 based 1T1R cell. (b) The *I*–*V* curves of reset operation for V_G ramp scheme. (c) The power generation during the RESET process for the two program schemes. (d) Cycle-to-cycle resistance distribution in 1T1R cell for two programming modes. Reproduced with permission from Ref. [114].

4.2. Endurance

Endurance, also called cyclability, refers to the ability of a memory cell to sustain switching cycles under repetitive SET/RESET operation with a predefined resistance window. The endurance requirement is different according to the intended application. As high as 10^{15} cycles in a non-stop condition is required for replacing DRAM-like storage class memory (SCM) applications, while 10^3 cycles would be enough for Flash memory application. In the 3D architecture, insufficient endurance could significantly decrease the write/read margin.

A three-layer 3D vertical $AlO_x/Ta_2O_{5-x}/TaO_y$ RRAM with extreme high endurance was demonstrated in Ref. [116]. The schematic view and cross-sectional TEM image of the architecture are shown in Figs. 12(a) and 12(b). The resistive switching layer is located at the sidewall between TE and BE vertically. The vertical devices show high layer-to-layer uniformity and similar electrical characteristics as the planar counterpart. The pulse conditions utilized during set and reset processes are −1.9 V, 100 ns and 2.3 V, 100 ns, respectively. As shown in Fig. 12(c), the devices in every layer exhibit 10^{10} switching endurance cycles. The degradation trend in vertical 3D devices, i.e., decreased HRS and stable LRS, are uniformity and show similar electrical characteristics as the planar counterpart. The retention results exhibit that both HRS and LRS could be kept more than 10^4 s at 125 °C without any degradation in the 3D architecture (Fig. 12(d)).

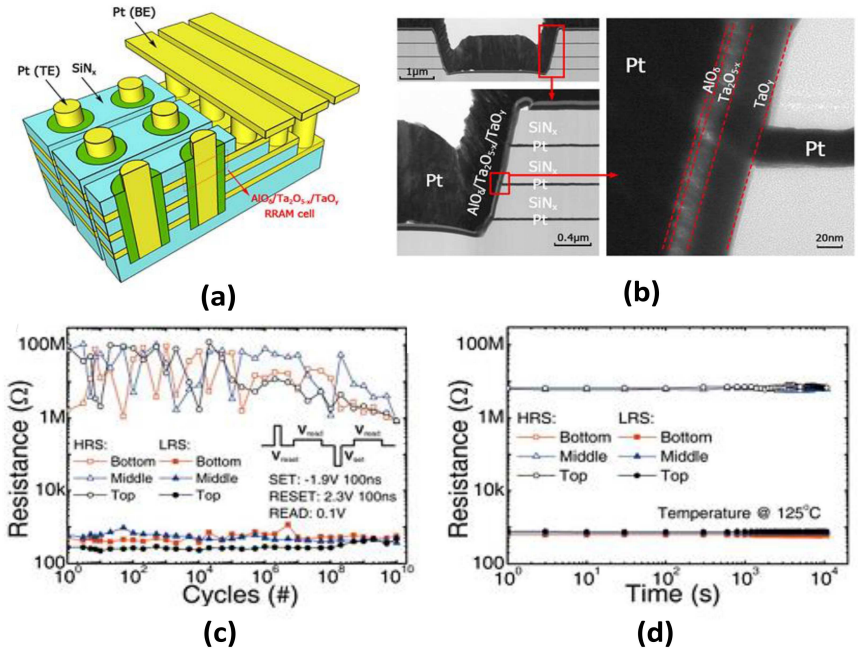


Fig. 12. (a) The schematic view of 3D vertical RRAM. (b) The cross-sectional TEM image of 3D vertical RRAM cells with the structure of $Pt/Ta_2O_{5-x}/TaO_y/Pt$. (c) All the devices demonstrate 1010 switching cycles. (d) The devices in triple layers show good retention. Reproduced with permission from Ref. [116].

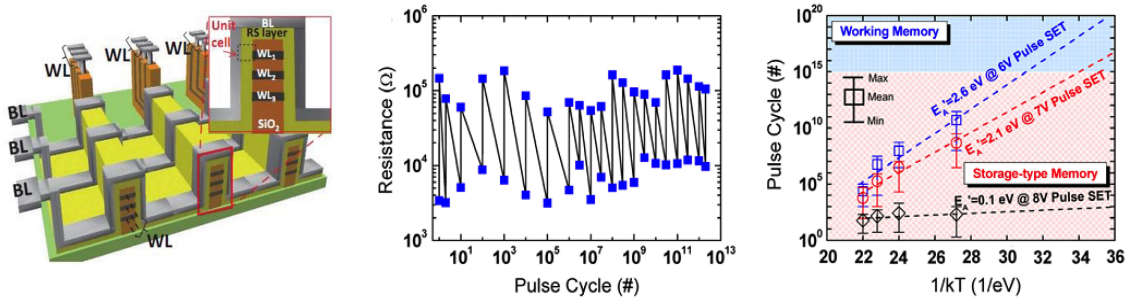


Fig. 13. (a) Illustration of 3D V-RRAM arrays. (b) Typical pulse endurance of over 1012 cycles achieved at room temperature. (c) Arrhenius plot of the endurance cycle as a function of the magnitude of SET pulses of 100 ns. A fixed RESET pulse of 100 ns at 6 V was used. Reproduced with permission from Ref. [117].

Hsu *et al.*[117] demonstrated an extrapolated endurance of over 10^{15} cycles at 100 °C for a Ta/TaO_x/TiO_2/Ti vertical RRAM with a sub-μA operating current, with schematic shown in Fig. 14(a) and room temperature endurance shown in Fig. 13(b). It is the best endurance that has been reported and can satisfy the requirements for replacing DRAM as working memory ($> 10^{15}$ cycles). They attributed the excellent endurance characteristics to the suppression of undesired overshoot current owing to the self-compliance and extremely low set current. The endurance would degrade to tens of cycles by increasing the set pulse to 8 V. The Arrhenius plot (Fig. 13(c)) of the endurance cycle elucidated that the irreversible hard breakdown was induced by the generation of new O_2 and V_O defects near the TaO_x/TiO_2 interface under electrical stress. This paper provided a good enlightenment to improve the endurance.

According to the current reports, the endurance characteristic is strongly affected by pulse voltage/duration, compliance level, activation energy, dielectric thickness, filament temperature, and filament size/shape. Engineering the device architecture, doping in the switching layer, and adopting novel read/write schemes have been proved as the effective methods to optimize the endurance characteristic.

4.3. Retention

A data retention longer than ten years at thermal stress up to 85 °C (operating temperature) and small electrical stress is expected for nonvolatile memory. The tail bits of the failure distribution become a limiting factor[118,119] for array level application. There are two types of retention failure: gradual degradation and abrupt degradation. For the retention failure with a gradual manner,[83,120–122] the resistance evolution can be extrapolated to find the failure point. However, a sudden resistance increase/decrease usually occurs rather than gradual transition during a typical retention process.[118,123,124] Then, an alternative method named temperature-accelerated method is adopted: varying the temperature/stress to record the time-to-failure/voltage-to-failure and draw the Arrhenius plot to extract the activation energy (E_a), and then extrapolate down to the operating temperature.

It is widely accepted that the retention loss is related to the diffusion of the V_O or metal particle (M). The gradual resistance degradation is an accumulation process. For LRS, The V_O/M concentration in filament is decreased by lateral diffusion from the surrounding region, resulting in degradation of the LRS retention characteristics. In the case of HRS, the vertical ions diffusion from the CF to the tunnel gap is a predominate parameter for the retention degradation. The abruptly retention failure phenomenon is correlated with a random process, which is often demonstrated by hopping-percolation model.[123] The cell resistance drastically increases when the most critical path is cut off during LRS retention measurement. HRS retention failure is due to reconstruction of the ruptured CFs due to the generation of V_O by thermal activation process or an electric field effect.

In principle, the retention is dependent on various factors, such as filament size,[121,125] filament number,[126] active energy,[127,128] tunnel gap length,[120] and oxygen vacancy/metal ions concentration.[118,119,123,129] The enhancement of the LRS retention can be achieved by taking advantage of the stronger filament, by applying a larger compliance current,[121,125] or novel programming scheme,[118,119,123,130] or material engineering.[127,128] By a two-step forming method,[118] the filament size can be scaled down while keeping high density of the oxygen vacancy, and the retention can be improved even under low-current operation. Employing the metal oxide electrolyte with higher activation energy for metal ions diffusion as the electrolyte, the diffusion probability of the metal ions would reduce in contrast to the ionic conducting electrolyte. Thus, the retention of RRAM can be improved significantly.[131] Bilayer structure is always adopted to improve the reliability of the RRAM,[127,128] different thermodynamics abilities of metal caps (Ti, Ta, Hf, etc.) to scavenge oxygen from HfO_2 are expected to result in different oxygen vacancy profiles. The absolute value of oxygen vacancies will be increased by using the metal cap, thus improving the retention properties. Moreover, the retention characteristics can be significantly improved by tuning the filament morphology through adjusting the operating conditions.[131] Bai[116] reported both HRS and LRS of each layer could be kept more

than 10^4 s at 125 °C without any degradation in three-layer 3D vertical $AlO_x/Ta_2O_{5-x}/TaO_y$ (Fig. 13(d)).

4.4. Distinguishable states

The distinguishable states of the memristor are also called the multi-level storage. Multi-level cell (MLC) capability characterizes the ability to store several bits in one cell. For general, the more the number of states, the better. It can greatly increase the storage density and computing power without increasing the cost. Currently, the maximum number of realizable states reported in the redox memristor can reach 64. According to the filamentary theory, the LRS resistance depends on the size of the conductive filament, while the HRS resistance lies on the gap length between filament and the electrode formed after the filament ruptured. There are two typical methods to realize the MLC operation through modulating the resistance states of RRAM.[8,90] During set process, the intermediate resistance states can be achieved by different set compliance. In 1T1R structure, MLC capability is highly depending on the controllability of the transistor. Xu *et al.*[135] revealed that the MLC margin would shrink due to the dynamic voltage drop on the transistor and memory cell. Controlling the RESET process is more favorable to achieve the uniform multilevel switching, which can be achieved both by changing the amplitudes or width of the RESET pulse.

For practical application, enough sense margins and tight uniformity of each resistance state are required for MLC operation.[116,133,134] The state fluctuation caused by stochastic nature of the conductive filament formation/rupture is a great challenge for the MLC capability.[83,113] The uniformity of the resistance states could be improved by optimizing the program and verify schemes.[133,135–138] Zhao *et al.*[133] proposed a pulse-training programming scheme to improve the uniformity of MLC without any read-verification operation. They adopted a train of pulses rather than a single pulse in RESET process to achieve better MLC control. By applying the novel scheme to a 3 bit/cell HfO_x-based RRAM device, the uniformity of resistance distribution was improved up to 80%. Yu[137] investigated the read disturb and bipolar read scheme on multilevel RRAM-based deep learning inference engine, proposed a bipolar read scheme to enhance the resilience against the read disturb. A self-limited operation scheme for improving the uniformity of multi-level resistance states was proposed in Ref. [115], demonstrated in a Ta_2O_5/TaO_x memristive structure shown in Fig. 14(a). Compared to conventional operation scheme (left panel of Fig. 14(b)), the self-limited operation scheme could diminish the excessive current during SET and avoid the over-RESET during switching, thus achieving better multilevel storage (right panel of Fig. 15(b)). The capability of MLC storage was also demonstrated in 3D vertically integrated RRAM cell.[84,116]

4.5. Scalability

Excellent scalability of RRAM cell has been demonstrated in several recent works. Joo *et al.*[139] theoretically studied the scaling limits of RRAM comprehensively based on atomic rearrangements calculation, by which the minimum number of atoms that need to be moved in order to create/rupture the conductive path with LRS current > 1 μA and ON/OFF ratio > 10 was estimated. A scaling limit of lateral size of 4 nm was suggested to grantee acceptable device performance and stability. By using Pt-coated silicon AFM tip as an electrode after directly contacting it with NiO surface, Lee *et al.*[14] successfully demonstrated an RRAM with ultra-small cell size (10 nm × 10 nm), which was approximately defined by the size of AFM tip. Tsai *et al.*[141] fabricated a sub-5 nm AlO_x based RRAM cell located at the intersection of the single-walled carbon nanotube (CNT) crossbar electrodes, the schematic and SEM are shown in Figs. 15(a) and 15(b).

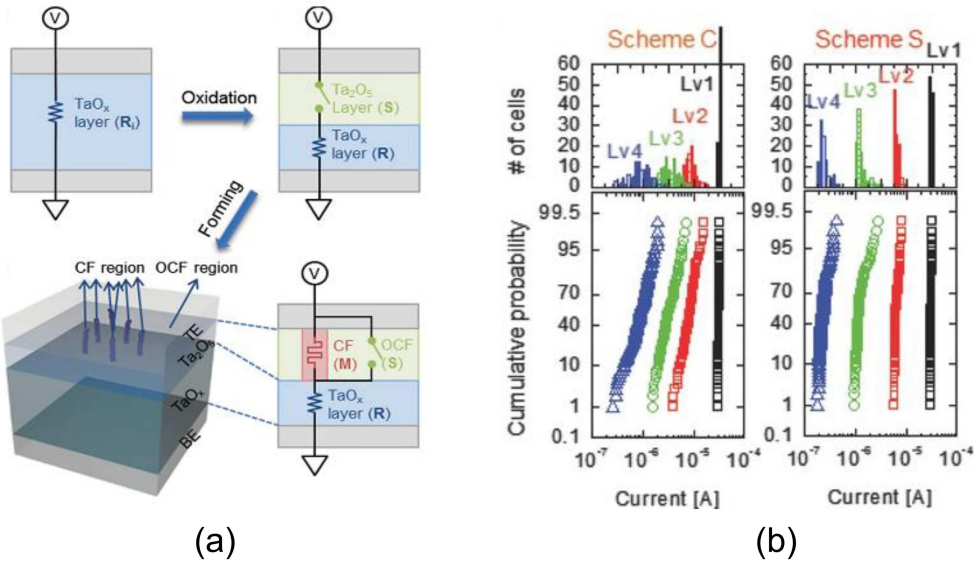


Fig. 14. Self-limited switching for uniformly regulating the values of both the low- and high-resistance states. (a) Self-limited operation scheme. (b) Improvement in the uniformity of multilevel switching due to self-limited switching. Left panel shows the multilevel switching uniformity of conventional operation scheme and right panel shows that of self-limited operation scheme. Reproduced with permission from Ref. [114].

Stable resistive switching characteristics with on/off ratio of 5×10^5 and sub μA operation current were achieved. The CNT resistance has a proportional relation with the LRS of RRAM cell, while they become independent as the CNT resistance gets below 10 MΩ (Fig. 15(c)). Both set and reset voltages of the RRAMs were inversely proportional to the CNT resistance (Fig. 15(d)). Impressive attempts to reduce cell size by using self-assembled oxide nanowire were made.[141–145] The fundamental scaling limits of RRAM device are greatly extended with the demonstration of the device with ultimate scaled size (1 nm–2 nm). The above-mentioned works present a promising potential of the scalable capability of RRAM device, however, an ultimate scaled device with fully integration into the CMOS process was still absent until the attempt from Govorean *et al.*[144] The world's smallest HfO_2-based RRAM cell with an area of less than 10 nm × 10 nm was presented. This device was fabricated by 65 nm CMOS process.[145] Fast ns range switching time at low voltage and ultralow switching energy < 0.1 pJ per bit were successfully demonstrated, as well as excellent endurance of more than 5×10^7 and large on/off ratio (> 50).

The prediction of the scaling limit of 3D RRAM array is much more complicated than that of a single device. As shown in Fig. 16(a), as the thickness of WL/insulator stack in 3D VRRAM scales continuously, the sneak leakage path will inevitably occur in the selective layer between adjacent WLs. Xu *et al.*[51] proposed an elegant building block of 3D VRRAM using self-aligned self-selective RRAM to solve this problem (Fig. 16(b). Figure 16(c) shows the device structure design of TiN/TiO_x/HfO_2/Ru SSC device. Obvious hysteresis loop could be observed in the typical *I*–*V* characteristics of this SSC device (Fig. 19(d)). Ultra-low off-state leakage current (pA) and operating current below μA were achieved. The high nonlinearity in LRS guarantees the enough capability to suppress the sneaking current in high density array. The SSC also exhibits excellent endurance and favorable thermal stability (Figs. 16(e)–16(f)). By successfully eliminating the inter-layer leakage, the scaling limit of VRRAM could be extended beyond 5 nm which was later confirmed by experiments.[96]

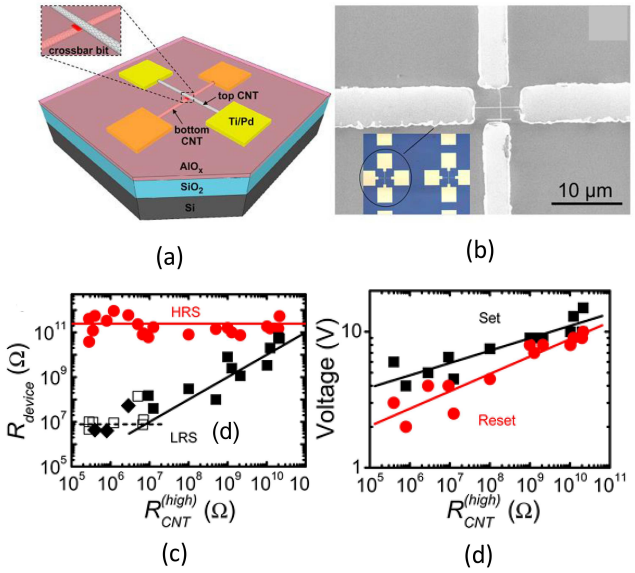


Fig. 15. (a) The AlO_x RRAM cell sized sub-5 nm at the intersection of carbon nanotube crossbar electrodes. (b) SEM of the RRAM bit structure, including metal contacts to CNTs. (c) The LRS of crossbar RRAM is independent of CNT electrode resistance near 10 MΩ. The HRS of crossbar RRAM is independent of the CNT electrode resistance. (d) the operation voltages of the crossbar RRAM is dependent on the more resistive CNT electrode resistance. Reproduced with permission from Ref. [141].

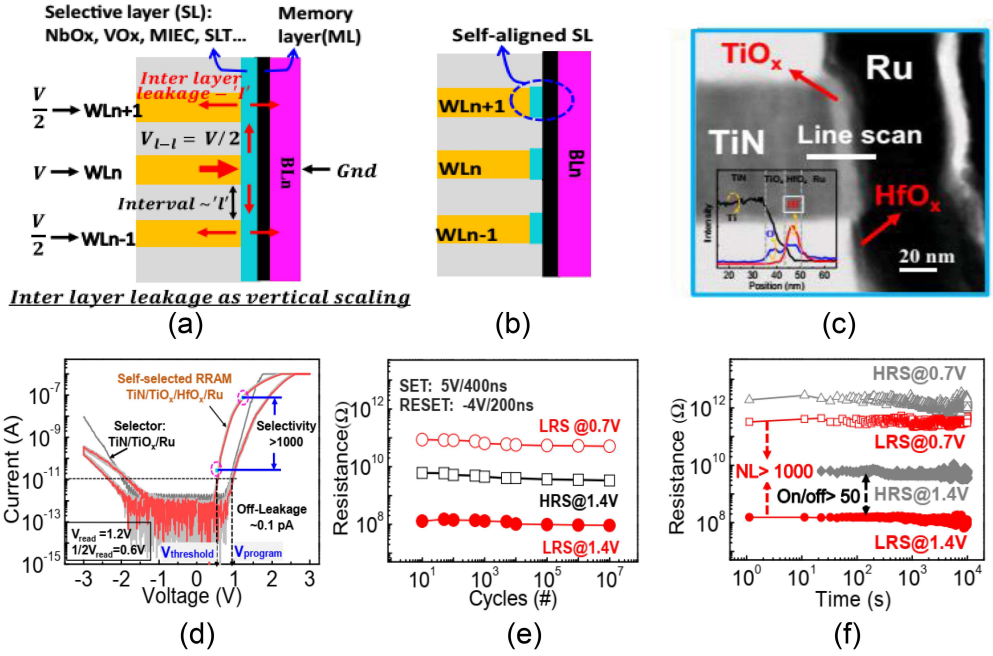


Fig. 16. (a) The inter-layer leakage problem in VRRAM occurs with the thickness scaling. (b) Schematic of the proposed self-aligned VRRAM. (c) TEM image of cross section of TiN/TiO_x/HfO_2/Ru SSC device. (d) Typical *I*–*V* curve of the SSC. (e) The proposed SSC for VRRAM can switch $> 10^7$ pulse cycles. (f) Both HRS and LRS of the SSC show favorable thermal stability. Reproduced with permission from Ref. [51].

5. Applications of 3D RRAM beyond storage

Resistive switching memory has potential application in the field of embedded circuit and high-density storage due to its excellent electrical properties, including high program speed, low consumption, and high density. More interestingly, the RRAM cells which show memristive switching behavior could be employed as synaptic devices for the neuromorphic and stochastic computing, which pave a new way to develop new computing system beyond Von Neumann architecture. The following section will focus on the topic of potential application of RRAM beyond storage.

5.1. Neuromorphic computing

Neuromorphic computation system which can imitate human's brain, is superior to the conventional computer in recognizing various objects and processing imprecise sensory inputs through the energy-efficient and fault-tolerant computation. The human brain consists of 10^{11} neurons and 10^{15} synapses,[146] the former executes computation by collecting information from the other neurons and generating a spike, while the latter facilitates the computation by adjusting the strength as a respond to neuronal activity, as shown in Fig. 17(a). The mimicry of biological neurons and synapses by electronic devices is the crucial challenge for the hardware implementation of such massively parallel, three-dimensional, compact systems. The first FET-based synapse was proposed by Carver Mead,[147] but this synapse includes a large number of transistors and consumes significant energy and hinders the scaling trend of the neuromorphic chips.

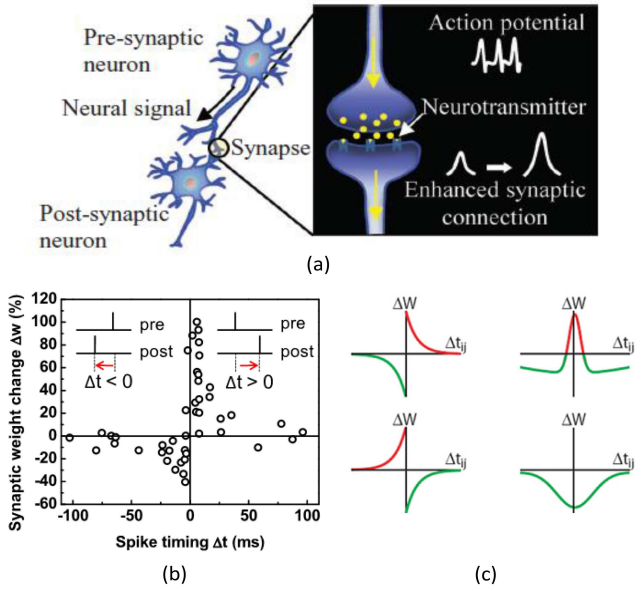


Fig. 17. (a) The weight change in synapse is determined by the relative timings of neuronal spikes from the pre-synaptic neuron and the post-synaptic neuron. (b) In STDP measurements in hippocampal glutamatergic synapses, the synaptic weight change is a function of relative timing of pre- and post-spikes. (c) Ideal STDP learning rules. Left: Asymmetric Hebbian learning rule and asymmetric anti-Hebbian learning rule. Right: Symmetric Hebbian learning rule and symmetric anti-Hebbian learning rule. Reproduced with permission from Ref. [164].

To achieve true human-level artificial intelligence, an electronic synapse with ultra-low power consumption, sub-10 nm scalability must be developed. To date, a broad spectrum of two-terminal non-volatile memory technologies with programmable conductance has been investigated to emulate synapses, such as phase change memory (PCM),[148–150] oxide based resistive switching memory (RRAM),[150–152] ferroelectric switches,[154,155] and carbon nanotube devices.[156–158] Among these candidates, RRAMs are more favorable due to their excellent scalability, fast switching speed, compatibility with CMOS technology and low energy consumption. The research activities in mimicking biological synapse by RRAM are growing rapidly since 2008.[151–153,159–163]

Activity-dependent synaptic plasticity is fundamental for learning and memory in neuronal systems, showing an analog-like behavior with numerous synaptic weight states modified by the neural activities in pre-synaptic and post-synaptic spikes.[164] The polarity or the degree of the synaptic weight change is influenced by the correlated activity of the presynaptic spike and postsynaptic firing, including the temporal relation, the spike rate, and the voltage amplitude.[151,153,164–167] Long-term potentiation (LTP) and long-term depression (LTD) are two opposite activity-dependent plasticity modifications, which are regarded as the bidirectional learning rule. According to the Hebb's theory, synaptic weight change is a function of the relative timing of pre- and post-synaptic spikes. This behavior calls spike-timing dependent plasticity (STDP) which is a critical function for learning and memory in our brains.[166] The case a pre-synaptic spike precedes a post-synaptic spike repeatedly contributes to the potentiation of the synapse, whereas, if a post-synaptic spike precedes a pre-synaptic spike repeatedly, the synapse depresses. Figure 17(b) demonstrates the asymmetric STDP adapting the original data from Bi and Poo's paper.[168] Figure 17(c) exhibits four typical forms of STDP which vary due to different spikes or different types of synapses, excitatory or inhibitory. Besides spike timing, the synaptic plasticity is greatly affected by the firing rate, spiking orders, and dendritic location. The spike-rate dependent plasticity (SRDP) is another central rule of learning, which reflects the influence of the synaptic activation frequency on the long-lasting modification. The synaptic efficacy is strengthened if the postsynaptic activity stays below a critical rate, and is conversely weakened under the high-frequency post-synaptic excitation. In addition, the LTP and LTD could be induced by repetitive coincident pre- and postsynaptic stimulations with low and high voltage amplitudes.

In the implementation of the synapse behavior, the terminals of the device are regarded as the pre- and post-synaptic sites, and the conductance/resistance of the device represents the synaptic weight or synaptic efficacy that refers to as the connection strength of the neurons. The more resis-

tance states, the better the capacity and robustness of the neuromorphic systems will be. To implement synaptic function, the RRAM device must show the continuous tuning of conductance/resistance by voltage pulse. Yu *et al.*[23] firstly reported metal oxide-based RRAM with a gradual resistance modulation by hundreds of identical pulses. This synaptic device consumes sub-pJ energy per spike which is $\times 1000$ reduction compared with the present-day PCM synaptic devices. They utilized the RESET process for adoptive learning since the RESET process is gradual and the SET process is abrupt. Li *et al.*[153] proposed the Ag/AgInSbTe/Ag structure which exhibits the reproducible gradual resistance tuning. Figures 18(a) and 18(b) demonstrate that the device conductance could continuously increase and decrease due to the consecutive positive or negative voltage sweep, which are also implemented under the pulse stimulation mode. The device conductance is equivalent to the synaptic weight. The synaptic weight can be modified in respond to a potentiating or depressing stimulus. Based on the memristive characteristic, the device was utilized to mimic the activity-dependent synaptic plasticity. The STDP learning rule was reproduced in the memoritive device by applying the paired spikes shown in Fig. 18(c).

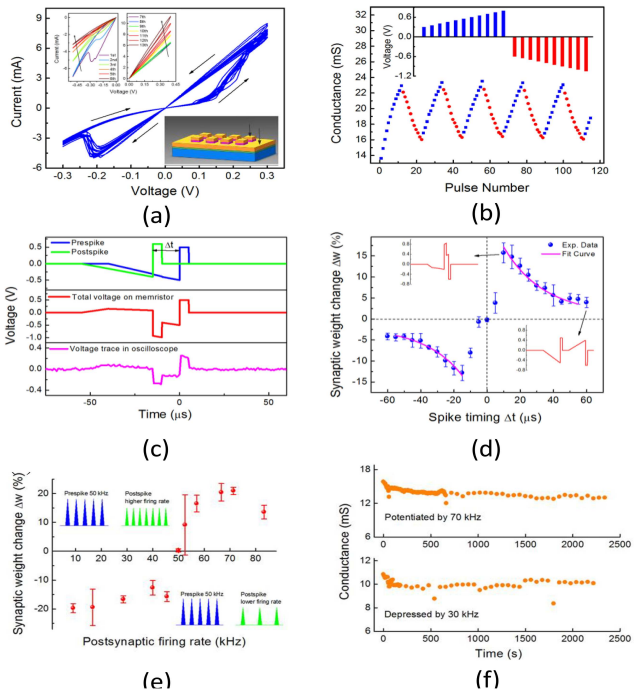


Fig. 18. (a) The *I*–*V* characteristics of the Ag/AgInSbTe/Ag device. Up inset: the device conductance continuously increases (decreases) with the consecutive positive (negative) voltage sweep. Lower inset: the simple stacked structure of the memristor. (b) Repetitive gradual conductance modulation under pulse stimulation mode. Inset: the pulse scheme applied on the device to achieve stable and repeatable conductance tuning. (c) Pulse scheme used to implement STDP. The upper shows a negative temporal difference that exists between the pre- and post-spikes. The middle shows the effective voltage on the memristor. The lower shows the voltage trace captured by an oscilloscope. (d) The asymmetric Hebbian learning STDP rule of the memristor. (e) Implementation of SRDP in the memristor. (f) The changes in synaptic conductance following potentiation (upper) and depression (lower) stimulations. The synaptic efficacy is a long-term modification. Reproduced with permission from Ref. [159].

The synaptic weight changes are plotted in Fig. 18(d) as a function of the time difference between the presynaptic spikes and the postsynaptic spikes.

Based the asymmetric Hebbian learning STDP rule, the other three biological STDP forms according to Fig. 18(c) are also demonstrated in this device. STDP has been implemented in many other RRAM systems, including HfO_x,[169] TiO_x,[17,171] Ag_2S,[159] AlO_x,[161] TaO_x,[172,173] IGZO,[174] Cu_2S,[175] and PEDOT:PSS.[176,177] In addition to the STDP, SRDP is also implemented in the Ag/AgInSbTe/Ag memristor. Figure 18(e) shows the dependence of synaptic modification on the postsynaptic firing rate induced by 50 repetitive stimulations.[153] The synapse is depressed if the postsynaptic firing rate is below a critical rate (50 kHz), and potentiated when the rate is beyond the threshold. Meanwhile, LTD could be induced by post-synaptic frequencies, whereas LTP results from the high frequencies. Both the LTP and LTD depend on the stimulation frequency and the synaptic voltage. As shown in Fig. 18(f), the synapse electric is potentiated and depressed by 70 kHz and 40 kHz, respectively. The nonvolatile property of the conductance further confirmed the long-term modification.

Extensive work has demonstrated that the brain-like associative learning, sequence learning, and pattern completion can be implemented with the synaptic devices.[151,160,162,164,176–180] Furthermore, computing applications with synaptic devices have focused on brain-inspired architectures and algorithms. Suri[162] proposed system-level simulations with binary CBRAM synapses by intrinsic probability programming methodologies. A 2-layer spiking neural network (SNN) is simulated for processing temporally video data. A video of cars passing on a freeway recorded with 128×128 pixel silicon retina is presented to the two-layered SNN consisting of about 2-million CBRAM synapses. The SNN shown here can be used as a prototype to realize more complex functions such as image classification, position detection, and target tracking. The same task of address-event-representation (AER) video data of driving in a car is widely demonstrated by RRAM based neural network.

The RRAM array shows excellent potential for the neuromorphic application with comparable areal density and power efficiency with the human cerebral cortex. However, the practical implementation of such array is very challenging because of the inevitable device variability. The current reports about memristor neuromorphic network were all based on disconnecting each device from the crossbar for individual forming or using an individual switch transistor at each cross-point.[181–183] These approaches conflict with the goal of achieving a neuromorphic network with extreme high density. Merrikh[183] reported an experimental demonstration of a fully operational neural network based on an integrated,

transistor-free crossbar with metal-oxide memristor. The uniformity and other important characteristics of the devices with Al_2O_3/TiO_{2-x} bilayer were optimized firstly, providing a foundation for the implementation. The SEM image of the memristor crossbar is shown in Fig. 19(a). A single-layer perceptron can successful be implemented by a 10×6 fragment of the memristive crossbar (Fig. 19(b)). Figures 19(c) and 19(d) show the pattern classification experiment results. If the output signal is equal to the correct class of the applied pattern, the classification is successful. After 23 epochs, the perfect classification was achieved with the standard deviation of ten epochs. This work is regarded as the primary step towards much larger and more complex memristive neuromorphic networks.

Recently, Yao *et al.*[184] demonstrated a complete five-layer convolutional neural network (CNN), one of the most important models for image recognition, implemented using a fully memristor-based hardware system. Based on the first integration system of multiple memristor crossbar arrays in the world, he successfully deployed a CNN on a highly consistent analog resistance array for the first time, and verified the prototype system of neural network processor based on memory integration. Figure 20(a) is the proposed memristor-based computing architecture for neural networks. The hardware system consists of 8 integrated PE arrays and other functional modules, which are implemented by the field-programmable gate array evaluation board (ZC706, Xilinx) and customized printed circuit board (PCB). Figure 20(b) presents the core PCB subsystem which consists of a memory integral unit based on the resistive memory array and the necessary functional modules implemented by peripheral discrete chips.

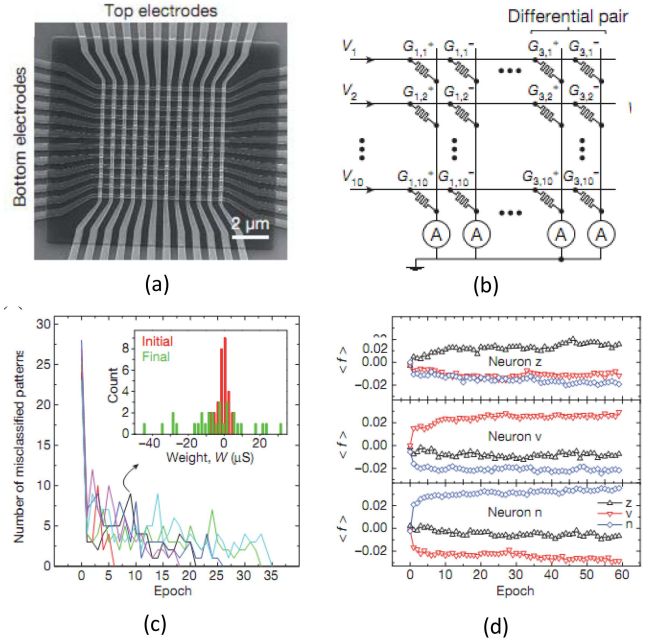


Fig. 19. (a) Integrated 12×12 crossbar with an Al_2O_3/TiO_{2-x} memristor at each cross-point. (b) Pattern classification experiment in physical-level description. The single-layer perceptron consists of 10×6 fragment of the memristive crossbar. (c) Pattern classification experiment results. After six training runs, different initial states were converged to the perfect value (zero). Insert: the distribution of weights in the initial state and after epoch 21. (d) The evolution of output signals, averaged over all patterns of a specific class. Reproduced with permission from Ref. [183].

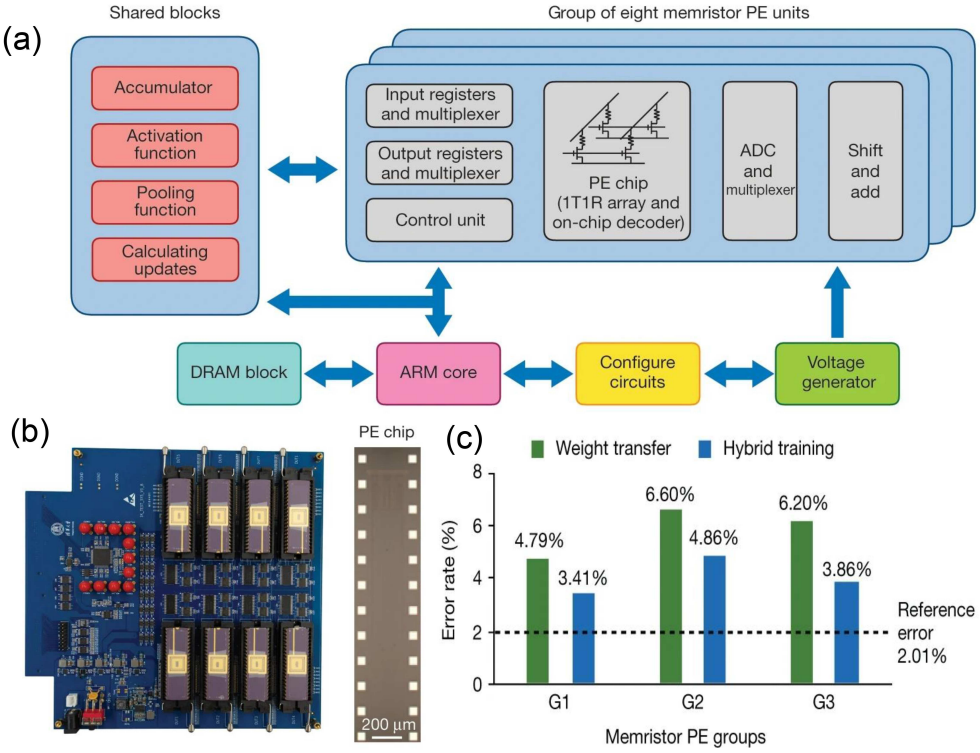


Fig. 20. Fully hardware-implemented memristor convolutional neural network. (a) Memristor-based hardware system: schematic of the system architecture with eight integrated memristor PE units and other functional blocks. (b) The photograph of the integrated PCB subsystem and the image of a partial PE chip consisting of a 2048-memristor array and on-chip decoder circuits. (c) The error rate obtained on the test set after hybrid training. Reproduced with permission from Ref. [184].

Yao innovatively proposed parallel hardware strategies for hybrid training and replicating the same convolution kernel into different arrays. This hardware strategy can improve the convolution operation efficiency through parallel acceleration while accommodating the non-ideal fluctuation and noise of the device, and effectively solve the key problem of realizing the CNN. The experimental results show that the recognition rate can be equivalent to that of the software when calculating the complete convolutional neural network, as shown in Fig. 20(c).

With the unremitting efforts of the researchers, neural networks could be fully implemented in RRAM based systems and executed on site. The ingenious neural networks could transform the computing paradigms from high-demand, sequential digital operations into an activation-driven, parallel analogue-computing network. However, the current memristor-based systems are all constructed on regular 3D crossbar arrays. In human brains, the synaptic network consisting of billions of neurons is an intricate 3D architecture. As such, it is very attractive to extend the 2D cross-point array to 3D array to improve the integration density.[185–191] For the first time, Wang *et al.*[185] developed a 3D synaptic architecture based on Ta/TaO_x/TiO_2/Ti RRAM with cost-efficiency, as shown in Fig. 21. The 3D synaptic device shows self-rectifying characteristics to overcome the sneak current problem through inactive synapses in the network (Fig. 21(b)). The STDP was successful implemented in the 3D synaptic device (Fig. 21(c)). The result shows a similar trend as that of a biological synapse. The ultra-low power consumption as low as 7 fJ/spike in 3D architecture is comparable to that in biological synapses (Fig. 21(d)), suggesting promising potential of the breakthrough hardware for future neuromorphic computation. The linearity of synaptic weight updates is defined as the linear response of cell conductance to consecutive potentiation or depression inputs. The linearity of synaptic devices is a critical parameter for learning accuracy of RRAM based hardware neural network. The author[186] further proposed a new state-independent bipolar-pulse-training scheme to improve the linearity of weight updates, as shown in Fig. 21(e). The calculated accuracy under different operating modes is illustrated in Fig. 21(f). The ideal linear update scheme achieved nearly 100% training accuracy in 60 cycles. Compared with the unipolar-pulse-scheme, the bipolar-pulse-training -scheme with a lower NL improved the training accuracy significantly.

Further, the neuromorphic hardware design based on the 3D VRRAM is discussed by researchers. The 3D VRRAM with the plane WL structure has been investigated for deploying machine learning algorithms due to bit-cost scalability and simple structure.[188–191] For the neuromorphic applications of 3D V-RRAM, the pillar electrodes are utilized as the input vector and the plane electrodes as the weighted-sum outputs. Such operation scheme limits the number of output neurons that equals to the number of vertical layers. Li *et al.*[188] proposed a novel operation scheme to combine the selected lines and the word-plane electrodes as input vector, and all the BLs are designed as weighted-sum outputs. Figures 22(a)–22(c) demonstrate the operation scheme of 3D V-RRAM toward feedforward/backward inference (read) mode and weight update (write) mode. Compared to the 2D implementation, the proposed 3D V-RRAM implementation shows larger write margin for weighted sum/weight update, smaller latency, and energy consumption for weight update. This work demonstrates the attractiveness for building a monolithic 3D neuromorphic hardware platform. Li[189] improved the energy–delay product of neuromorphic computing by building the 3D vertical RRAM integrated with FinFET. Figure 22(d) exhibits the structure diagram. The 3D vertical RRAM is used for the system-level simulations as shown in Fig. 22(e). In the unsupervised winner-take-all visual system, the 3D architecture 'folds' neurons/synapses into balanced plane with dense connections. The interconnect RC effects and long sneak leakage paths are thus reduced. As shown in Fig. 22(f), the 3D architecture improves the energy–delay product by 55% and reduces V_{DD} by 74%, which is a big improvement over the conventional 2D architecture.

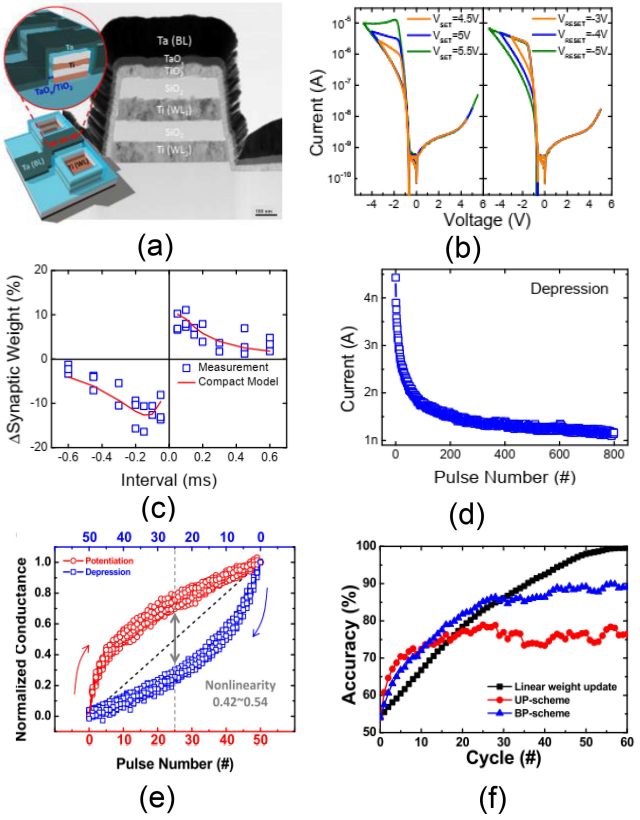


Fig. 21. (a) Cross-sectional TEM image of 3D Ta/TaO_x/TiO_2/Ti double-layer V-RRAM as 3D synaptic network. (b) Multiple resistance states can be achieved by varying V_{SET} or V_{RESET}. (c) STDP measurement result in the 3D Ta/TaO_x/TiO_2/Ti device. (d) Depressing characteristics in the 3D device. The energy consumption per spike is as low as 7 fJ. (e) Normalized synaptic characteristics is optimized by the proposed the bipolar-pulse-training scheme. (f) Training accuracy with the training cycle under different training scheme. Reproduced with permission from Refs. [185,186].

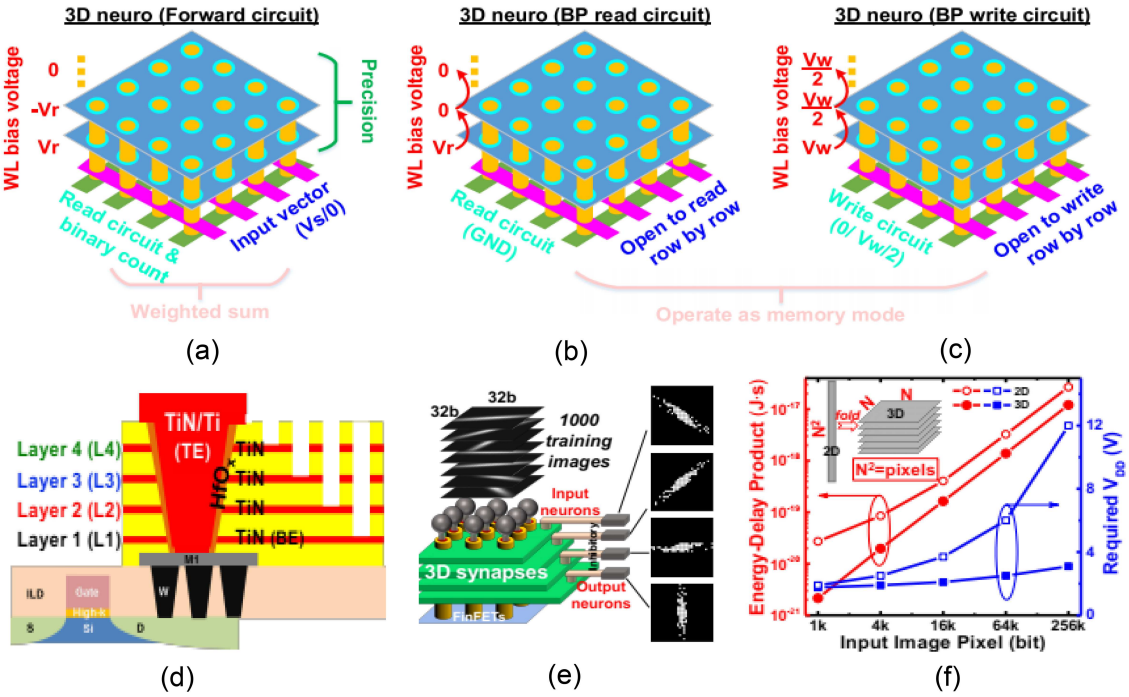


Fig. 22. Operation mode of 3D V-RRAM array for neuromorphic computing. (a) Feedforward/backward scheme for propagation mode to accelerate the weighted sum. (b) Read scheme and (c) write scheme for weight-update mode. (d) Schematic of 4-layer 3D RRAM integrated in FinFET platform. (e) A 3D neuromorphic visual system based on a $32 \times 32 \times 4$ 3D array. A winner-take-all network with stochastic learning rule is simulated for orientation detection. (f) SPICE-simulated energy–delay product and required V_{DD} to program worst case-located synapses in 2D and 3D WTA networks. Reproduced with permission from Refs. [188,189].

Since the 3D VRRAM with even/odd WL structure can contain doubled cell bits compared to the WL plane structure, it exhibits outstanding area efficiency and high potential in neural network implementation. However, the increased interconnect lines due to the split WLs require a sophisticated control mechanism to access the cells and exacerbate the sneak path issue. To minimize the side effects induced by the increase of interconnects, Kim *et al.*[191] proposed the hardware design through a more balanced structure along the three orthogonal directions. In this design, the input data was fed through the vertical pillar electrodes and the output was detected at the WLs. By this approach, a more balanced array configuration on layers, i.e., similar row and column lengths, was achieved and the elongated sneak path was reduced efficiently. Furthermore, they defined a weight adjustment method to harmonize with the use of even/odd lines and devised a sequential operation to ensure to program only intended cells. The SSC with high nonlinearity was also utilized to alleviate the sneak path issue. This work verifies the effectiveness of the even/odd WL structure for neuromorphic system implementation.

As neural networks have been rapidly evolving, 3D vertical RRAM is considered as the most promising candidate to increase the density of neuromorphic circuits. Although both of 3D VRRAMs organized in the plane WL structure and the even/odd WL structure show potential applications. The complicated interconnection of 3D VRRAM makes the design more challenging, mainly due to several issues, such as the vulnerable unselected cells that share the same lines with the selected cells are more easily to be affected and the sneak leakage path per area increases. So, 3D VRRAM requires more carefully designed operations before building a monolithic 3D neuromorphic hardware platform.

5.2. 3D CMOL circuit

Hybrid CMOS/molecular circuits (CMOL) are a promising hybrid structure that extends the Moore's law 10 to 15 years beyond the so-called "redbrick wall".[192–194,194] Figure 23 demonstrates the basic idea of CMOL circuits, which corporate the CMOS subsystem with nanowire crossbars and simple two-terminal molecular devices (Figs. 23(a) and 23(b)).[195] The CMOS subsystem exhibits the advantages of flexibility and high fabrication yield. The typical *I*–*V* curve of the two-terminal molecular device is shown in Fig. 23(c). For the hybrid circuits, the efficient implementation of the integration of CMOS and nano crossbars is the main issue. This challenge can be overcome by area-distributed interface between the two elements, using conic-shaped vertical plugs "pins" in 3D CMOL. Because of the self-assembly of the molecular devices on the nanowire crossbar, the fabrication of the CMOL is much cost effective without precise alignment. Furthermore, the density of active devices in CMOL circuits is estimated up to $\sim 10^{12}$ active devices per cm^2 since the minimum feature of the CMOL is defined by the size of a certain molecule rather than the traditional lithography.[194] Borghetti *et al.*[196] achieved a 100-gate-scale hybrid CMOS/memristor circuit, in which memristive TiO_{2-x} film is integrated onto a foundry-built CMOS platform. An area- distributed CMOL-like interface with tilted crossbar was fabricated with nanoimprint technology.

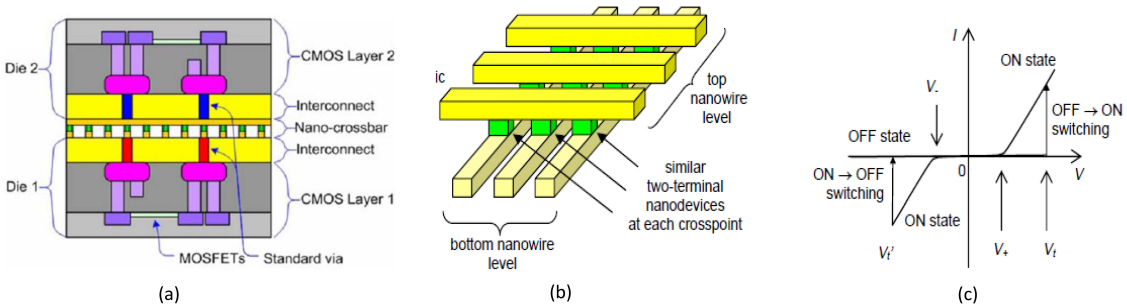


Fig. 23. (a) Hybrid circuit cartoon, (b) crossbar topology, and (c) micrograph of array of the memristive devices, and typical switching *I*–*V* curves. Reproduced with permission from Refs. [195,198].

The most fundamental application of the 3D CMOL circuit is the memory category, considering as the extension of the RRAM.[194,196,198,199] For such memory, each memristive device plays the role of a single-bit memory cell, while the CMOS subsystem works for the peripheral including input/output, coding/decoding, line driving, and sense amplification. The hybrid CMOS/memristor circuits could potentially combine all the advantages of the CMOS chip and the RRAM – high yields, cost efficient, high speed, low energy, and high endurance of the memory. Zhong and Chen[200,201] have taken the first experimental steps to implement the hybrid memories.

The high function density is one of the attractions of the CMOL memory, which can be approached by larger crossbar array. As the array size increases, the readout delays and voltage drop across crossbar wires become longer, as well as the leakage current via the memristor. When combined in circuits, the memory cells are organized in roughly square arrays with multiple cells sharing row (word) and column (bit) lines. The integration of a traditional transistor with the memory cell can enable unique access (read/write) to the row of cells at a time via bit lines by asserting the corresponding select word line (Fig. 24(a)). However, the peripheral area overhead (i.e., sense amplifiers, decoders, etc.) is proportional to $N \times \log N$ for the $N \times N$ crossbar array, which causes large area waste.[193,194] The passive crossbar is a much denser alternation, in which the select functionality is implemented via the diode-like *I*–*V* of the cross-point device (Figs. 24(b) and 24(c)). Therefore, the strong nonlinearity in the *I*–*V* is the expectation for the RRAM in the context of passive crossbar memories.

Novel defect and fault tolerance schemes must be considered for the CMOL memory. Two major techniques, the memory matrix reconfiguration and error correction code techniques (ECC), are always applied for increasing the defect-tolerance.[202,203] Memory matrix reconfiguration means replacing memory array lines containing bad cells by spare lines. Dmitri *et al.*[198] demonstrated that by using the best (but hardly practicable) reconfiguration and block size optimization, hybrid memories may achieve defect tolerance up to $\sim 15\%$. It is said that CMOL memories may be the first technology to reach the terabit frontier.

The most important application of CMOL technology may be in reconfigurable Boolean logic circuits[24,199,204,205] whose structure resembles the cell-based field programmable gate arrays (FPGAs). Figure 25(a) shows the schematic of original CMOL FPGA. The basic cell consists of an inverter and two pass transistors that serve two pins as the cell input and output. The cells connect to the nanowire/molecular subsystem via two pins (Fig. 25(b)). Each pin of the cell can be connected through a nanowire–nanodevice–nanowire link within a square-shaped "connectivity domain" around the pin when turning programmable diode. The RRAM here provides resistance states as the Boolean logic stage, and allows circuit mapping on CMOL fabric and its reconfiguration around defective nanodevices. Any logic function can be implemented by the logic gates (e.g., NOR, NAND) by CMOL-FPGA.

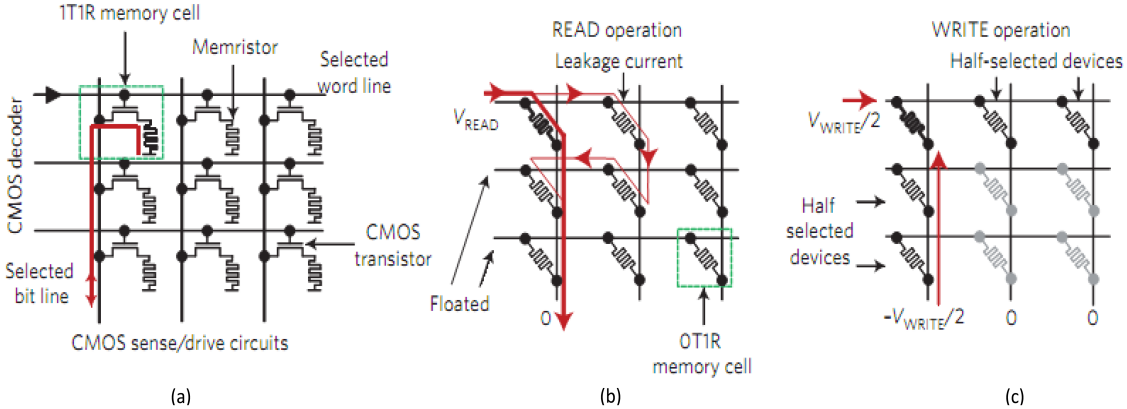


Fig. 24. The integrated architecture. (a) The active array. A 1T1R cell architecture enables unique access (read/write) to the cells in a selected row at a time via bit lines by asserting the corresponding select word line. (b) The passive crossbar memory. The select functionality is implemented via the diode-like *I*–*V* of the cross-point device. (c) Strong nonlinearity in switching dynamics helps avoid another problem specific to 0T1R architecture. Reproduced with permission from Ref. [197].

The CMOL FPGA circuits could improve the density by two orders of magnitude over purely CMOS FPGA circuits since the configuration information can be stored by the memristor in CMOL FPGA circuits while the implication of the configuration information in traditional FPGAs typically consumes 50%–90% of the chip area.[199] This advantage is confirmed in Ref. [24], where the CMOL FPGA circuits showed significant advantage of area-efficiency with manageable power consumption and high defect tolerance (above 20% of bad nanodevices). Xia *et al.*[204] implemented the first CMOL FPGA circuit by integrating the TiO_x memristor crossbar on top of a CMOS substrate using nanoimprint lithography, as shown in Fig. 25(c). The size of each cell in this 8×8 array is 50 μm × 50 μm. The logic part provides three types of computing elements: logic gates, flip-flops, and repeaters. By configuring the memristors properly, logic functions including OR, AND, NAND, NOR, and a positive-edge triggered D flip-flop can be implemented in the hybrid circuits. The equivalent computing circuits and the correspondingly measured truth tables for the logic gates of the hybrid circuits are shown in Fig. 25(d).

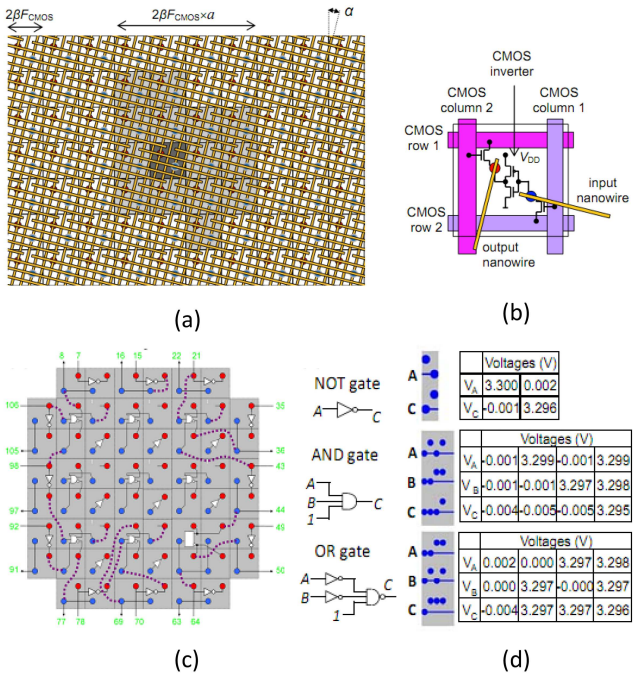


Fig. 25. (a) The original CMOL, (b) a single CMOS cell. (c) A successful experimental demonstration of the CMOL FPGA: CMOS layer fabric on a die and possible wiring for digital circuits using memristors. (d) The equivalent computing circuits, visualized digital results, and the measured truth tables for the logic gates in the hybrid circuits. Reproduced with permission from Ref. [206].

The properly designed CMOL circuits show high defect-tolerant, allowing ∼ 10% of bad bits in memories,[203] ∼ 20% for FPGA-like logic,[199] and more than 40% for some neuromorphic circuits.[207] The most exciting application of CMOL circuits is emulating the neuromorphic networks ("CrossNets").[9,204,208] In these circuits, the neural cell bodies ("somas") are implemented in the CMOS subsystem, and axons and dendrites are emulated by crossbar nanowires. This application concept has been comprehensively addressed in the previous section.

6. Conclusion and perspectives

As an advanced extension of two-dimensional passive crossbar array, the three-dimensional architecture should not only overcome the common problems that the passive arrays face, such as sneaking current path, voltage drop on line resistance, scaling, etc, but also need to deal with more challenging technological issues, such as complexity of fabrication process, thermal crosstalk, variation, etc. In order to suppress the sneaking current paths, each cell at the cross-point node should have nonlinear characteristics, which could be realized by explicitly introducing separate selector device or implicitly making the memory device self-selective. A variety of selector devices have been developed in the past few years, such as MIT or threshold switching device, bidirectional Schottky diode, and nonlinear tunneling device. Although great progresses have been made on the selector device with respect to large nonlinearity and high current density, additional attention should be paid on the aspects of CMOS compatible materials developing, off-state leakage reduction, and voltage window increment.

Promising cross-bar array designs featured 1S1R and 1D1R were already demonstrated as prototypical 3D stacked horizontal ReRAM (H-RRAM) arrays. Therefore, the stacked planar crossbar array at the current stage becomes the major choice for test chip designing and more readily for fabrication with mature CMOS platform. The recent demonstrated high volume 3D RRAM test chips (Panasonic's 8 Mb chip and Sandisk's 32 Gb chip) are all based on stacked planar crossbar. However, the cost advantage of VRRAM architecture over stacked array is competitive, which urges the research community making continuous effect to develop satisfactory SSC cell. The strategy of designing an SSC device should eliminate the secondary effect of the memory switching on its selectivity as much as possible. It is more difficult for self-selected device to tune its nonlinearity and memory performance separately considering the motions of oxygen or metal ions during memory operation may inevitably influence the barrier properties contributing to the nonlinearity. In this sense, the hybrid structure with dedicated threshold switching layer and memory switching layer should be a good choice. More efforts need to be made in the direction of avoiding interface diffusion with no material contamination, reducing the switching current with less thermal effect, and enhancing the reliability after continuous reading and writing.

For practical application, the main target of 3D RRAM is for massive data storage, high endurance as 10^{15} for DRAM is not necessary. This type of memory should not replace the flash because of the mature fabrication technology and improved performance for the traditional flash memory. It may be used as a complementary component in the memory hi-

erarchy, similar to the X-point. The 3D RRAM would improve overall system performance by serving as a non-volatile cache/buffer memory between a storage and the processor.

Beyond mere data storage, the RRAM technology also offers other applications, such as neuromorphic computing. The high similarity between the RRAM device and biological synapse makes the RRAM devices possibly serve as electric synapses to build a brain inspired frontier neuro computing system. Since 2008, the research activities in this area have been growing dramatically. A variety of synaptic functions such as SDTP, SRTP, LTP, LDP, etc have been demonstrated. Although the 3D RRAM array shows excellent potential for the neuromorphic application with comparable areal density and power efficiency with the human brain, the practical implementation is very challenging because of the complicated interconnection of 3D VRRAM. The device variability and the sneak current should be well addressed. Most of the system level demonstrations were only performed by simulation. Much research efforts are still needed to explore the resistive switching device with better uniformity and lower power consumption. The 3D VRRAM requires more carefully design before building a 3D neuromorphic hardware platform.

References

[1] Taur Y 1999 *IEEE Spectrum* **36** 25

[2] Tanaka H, Kido M, Yahashi K, Oomura M, Katsumata R and Kito M 2007 *Symposium on VLSI Technology*, June 12–14, 2007, Kyoto, Japan, p. 14

[3] Katsumata R, Kito M, Fukuzumi Y, Kido M, Tanaka H and Komori Y 2009 *Symposium on VLSI Technology*, June 15–17, 2009, Kyoto, Japan, p. 136

[4] Jang J, Kim H S, Cho W, Cho H, Kim J and Shim S I 2009 *Symposium on VLSI Technology*, June 15–17, 2009, Kyoto, Japan, p. 136

[5] Whang S, Lee K, Shin D, Kim B, Kim M and Bin J 2010 *IEEE International Electron Devices Meeting (IEDM)*, December 6–8, 2010, San Francisco, CA, USA, p. 29.7.1

[6] Waser R, Dittmann R, Staikov G and Szot K 2009 *Adv. Mater.* **21** 2632

[7] Marchewka A, Roesgen B, Skaja K, Du H, Jia C L, Mayer J, Rana V, Waser R and Menzel S 2016 *Advanced Electronic Materials* **2** 1500233

[8] Niraula D and Karpovb V 2018 *J. Appl. Phys.* **124** 174502

[9] Yang J J, Strukov D B and Stewart D R 2013 *Nature Nanotechnology* **8** 13

[10] Jacobs-Gedrim R B, Hughart D R, Agarwal S, Vizkelethy G and Marinella M J 2019 *IEEE Transactions on Nuclear Science* **66** 54

[11] Lv H B, Wan H and Tang T 2010 *IEEE Electron Device Lett.* **31** 978

[12] Liu Q, Sun J, Lv H B, Long S B, Yin K B and Wan N 2012 *Adv. Mater.* **24** 1844

[13] Chen H Y, Brivio S, Chang C C, Frascaroli J, Hou T H, Hudec B, Liu M, Lv H B, Molas G, Sohn J, Spiga S, Teja V M, Vianello E and Philip Wong H S 2017 *Journal of Electroceramics* **39** 21

[14] http://www.itrs.net/Links/2013ITRS/Home2013.htm

[15] Chen C B, Xu X X, Li X Y and Li Y T 2018 *Semiconductor Technology* **5** 1

[16] Pang Y, Wu H Q, Gao B, Ning D, Dong W, Rui L, Yu S M, Chen A and Qian H 2017 *IEEE Electron Device Lett.* **38** 168

[17] Wang M, Luo W, Wang Y, Yang L, Zhu W and Zhou P 2010 *Symposium on VLSI Technology (VLSIT)*, June 15–17, 2010, Honolulu, HI, USA, p. 89

[18] Zheng H X, Chen M C, Chang T C, Su Y T and Sze S M 2019 *IEEE Transactions on Electron Devices* **66** 4706

[19] Mei C Y, Shen W C, Wu C H, Chih Y D, King Y C and Lin C J 2013 *IEEE Electron Device Lett.* **34** 1253

[20] Rahaman S Z, Lee H Y, Chen Y S, Lin Y D, Chen P S, Chen W S, Chen W S and Wang P H 2017 *Appl. Phys. Lett.* **110** 213501

[21] Zhang J, Ding Y Q, Xue X Y, Gang J, Wu Y X and Xie Y F 2010 *IEICE Transactions on Electronics* **93** 1692

[22] Zhang Q, Cui X, Xu X, Wang X A and Zhou S 2016 *IEEE International Conference on Electron Devices and Solid-State Circuits (EDSSC)*, August 3–5, 2016, Hong Kong, China, p. 226

[23] Yu S M, Gao B, Fang Z, Yu H, Kang J and Wong H S P 2012 *IEEE International Electron Devices Meeting (IEDM)*, December 10–13, 2012, San Francisco, CA, USA, p. 10.4.1

[24] Madhavan A, Sherwood T and Strukov D B 2018 *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 26, p. 2759

[25] Kim G H, Kim K M, Seok J Y, Lee H J, Cho D Y and Han J H 2010 *Nanotechnology* **21** 385202

[26] Shrestha P R, Nminibapiel D M, Campbell J P, Ryan, J T, Veksler D, Baumgart H and Cheung K P 2018 *IEEE Transactions on Electron Devices* **65** 108

[27] Saraswat V, Prasad S, Khanna A, Wagh A and Ganguly U 2020 *IEEE Transactions on Electron Devices* **67** 3610

[28] Chand U, Huang K C, Huang C Y and Tseng T Y 2015 *IEEE Transactions on Electron Devices* **62** 3665

[29] Kim S, Zhou J, and Lu W D 2014 *IEEE Transactions on Electron Devices* **61** 2820

[30] Jin Y J, Xu Z, Yoon S F, Chia C K, Wang S J and Chi D Z 2016 *Electronic Materials Letters* **12** 365

[31] Kim T W, Zeigler D F, Acton O, Yip H L, Ma H and Jen A K Y 2012 *Adv. Mater.* **24** 828

[32] Lee M J, Seo S, Kim D C, Ahn S E, Seo D H and Yoo I K 2007 *Adv. Mater.* **19** 73

[33] Wang G, Lauchner A C, Lin J, Natelson D, Palem K V and Tour J M 2013 *Adv. Mater.* **25** 4789

[34] Luo Q, Xu X, Lv H, Gong T, Long S, Liu Q, Sun H, Li L, Lu N and Liu M 2016 *IEEE International Electron Devices Meeting*, December 3–6, 2016, San Francisco, CA, USA 2016, p. 11.7.1

[35] Seok J Y, Song S J, Yoon J H, Yoon K J, Park T H and Kwon D E 2014 *Advanced Functional Materials* **24** 5316

[36] Zhang L, Cosemans S, Wouters D J, Groeseneken G, Jurczak M and Govoreanu B 2014 *IEEE 6th International Memory Workshop (IMW)*, May 18–21, 2014, Taipei, Taiwan, China, p. 1

[37] Chen A 2013 *IEEE International Electron Devices Meeting (IEDM)*, December 9–11, 2013, Washington, DC, USA, p. 30.3.1

[38] Deng Y, Chen H Y, Gao B, Yu S, Wu S C and Zhao L 2013 *IEEE International Electron Devices Meeting (IEDM)*, December 9–11, 2013, Washington, DC, USA, p. 25.7.1

[39] Sun R, Chen H, Wang G R, Wang C and Hao L 2019 *Semiconductor Science and Technology* **34** 115023

[40] Burr G W, Shenoy R S, Virwani K, Narayanan P, Padilla A and Kurdi B 2014 *Journal of Vacuum Science & Technology B* **32** 040802

[41] Russo U, Ielmini D, Cagli C and Lacaita A L 2009 *IEEE Transactions on Electron Devices* **56** 193

[42] Mickel P R, Lohn A J, James C D and Marinella M J 2014 *Adv. Mater.* **26** 4486

[43] Sun P X, Li L, Lu N D, Li Y T, Wang M and Xie H W 2014 *Journal of Computational Electronics* **13** 432

[44] Ielmini D 2011 *IEEE Transactions on Electron Devices* **58** 4309

[45] Raghavan N 2016 *Microelectronics Reliability* **64** 54

[46] Baek I, Kim D, Lee M, Kim H J, Yim E and Lee M 2005 *IEEE International Electron Devices Meeting Technical Digest.*, December 5–7, 2013, Washington, DC, USA, p. 750

[47] Sharath S U, Vogel S, Molina-Luna L, Hildebrandt E, Wenger C, Kurian J, Duerrschnabel M, Niermann T, Niu G, Calka P, Lehmann M, Kleebe H J, Schroeder T and Alff L 2017 *Advanced Functional Materials* **27** 1700432

[48] Park S G, Yang M K, H Ju, Seong D J, Lee J M and Kim E 2012 *IEEE International Electron Devices Meeting (IEDM)*, December 10–13 2012, San Francisco, CA, USA, p. 20.8.1

[49] Banerjee W and Hwang H 2020 *Advanced Electronic Materials* **6** 9

[50] Chai Z, Shao W, Zhang W D, Brown J, Degraeve R, Salim F D, Clima S, Hatem F, Zhang J F, Freitas P, Marsland J, Fantini A, Garbin D, Goux L and Kar G S 2020 *IEEE Electron Device Lett.* **41** 228

[51] Xu X X, Luo Q, Gong T C, Lv H B and Liu M 2016 *Symposium on VLSI Technology (VLSIT)*, June 12–14, 2013, Honolulu, HI, USA, p. 1

[52] Gong T C, Luo Q, Lv H B, Xu X X, Yu, J, Yuan P, Dong D N, Chen C B, Yin J H, Tai L, Zhu X, Liu Q, Long S B and Liu M 2018 *IEEE Electron Device Lett.* **39** 1152

[53] Son M, Lee J, Park J, Shin J, Choi G and Jung S 2011 *IEEE Electron Device Lett.* **32** 1579

[54] Cha E, Woo J, Lee D, Lee S, Song J and Koo Y 2013 *IEEE International Electron Devices Meeting (IEDM)*, December 9–11, 2013, Washington, DC, USA, p. 10.5.1

[55] Lee W, Park J, Shin J, Woo J, Kim S and Choi G 2012 *Symposium on VLSI Technology (VLSIT)*, June 12–14, 2012, Honolulu, HI, USA, p. 37

[56] Lee W, Park J, Kim S, Woo J, Shin J and Choi G 2012 *ACS Nano* **6** 8166

[57] Song Y H, Park S Y, Lee J M, Yang H J and Kil G H 2011 *IEEE Electron Device Lett.* **32** 1023

[58] Kil G H, Yang H J, Lee G H, Lee S H and Song Y H 2012 *Jpn. J. Appl. Phys.* **51** 04DJ02

[59] Lashkare S, Karkare P, Bafna P, Raju M, Srinivasan V and Lodha S 2013 *IEEE International Memory Workshop (IMW)*, May 26–29, Monterey, CA, USA, p. 178

[60] Lee M, Kao C Y, Yang C L, Chen Y, Lee H and Chen F 2011 *69th Annual Device Research Conference (DRC)*, June 20–11, 2011, Santa Barbara, CA, USA, p. 89

[61] Lee M J, Park Y, Kang B S, Ahn S E, Lee C and Kim K 2007 *IEEE International Electron Devices Meeting(IEDM)*, December 10–12, 2007, Washington, DC, USA, p. 771

[62] Kang B S, Ahn S E, Lee M J, Stefanovich G, Kim K H and Xian W X 2008 *Adv. Mater.* **20** 3066

[63] Fu L P, Chen S K, Wu Z W, Li X Y, You M Y, Fan X L, Gao X P and Li Y T 2020 *Mod. Phys. Lett. B* **34** 12

[64] Anbarasu M, Wimmer M, Bruns G, Salinga M and Wuttig M 2012 *Appl. Phys. Lett.* **100** 143505

[65] Lopez J M, Robayo D A, Grenouillet L, Carabasse C and Molas G 2020 *IEEE International Memory Workshop (IMW)*, 17–20 May 2020, Dresden, Germany, p. 1

[66] Shenoy R, Gopalakrishnan K, Jackson B, Virwani K, Burr G and Rettner C 2011 *Symposium on VLSI Technology (VLSIT)*, June 14–15, 2011, Kyoto, Japan, p. 94

[67] Virwani K, Burr G W, Shenoy R S, Rettner C T, Padilla A and Topuria T 2012 *IEEE International Electron Devices Meeting (IEDM)*, December 10–13, 2012, San Francisco, CA, USA, p. 2.7.1

[68] Burr G W, Virwani K, Shenoy R S, Fraczak G, Rettner C T and Padilla A 2013 *Symposium on VLSI Technology (VLSIT)*, June 11–13 Kyoto, Japan, p. T66

[69] Jo S H, Kumar T, Narayanan S, Lu W D and Nazarian H 2014 *IEEE International Electron Devices Meeting (IEDM)*, December 15–17, 2014, San Francisco, CA, USA, p. 6.7.1

[70] Lee, Lee D, Woo J, Cha E, Park J and Hwang H 2014 *IEEE Electron Device Lett.* **35** 1022

[71] Luo Q, Xu X X, Liu H T, Lv H B, Gong T C, Long S B and Liu Q 2015 *IEEE International Electron Devices Meeting (IEDM)*, December 7–9, 2015, Washington, DC, USA, p. 10.4.1

[72] Zhao X L, Ma J, Xiao X H, Liu Q Shao L, Chen D, Liu S, Niu J B, Zhang X M, Wang Y, Cao R R, Wang W, Di Z F, Lv H B, Long S B and Liu M 2018 *Adv. Mater.* **30** 1705193

[73] Kim S, Liu X J, Park J, Jung S, Lee W and Woo J 2012 *Symposium on VLSI Technology (VLSIT)*, June 12–14, 2012, Honolulu, HI, USA, p. 155

[74] Luo Q, Xu X X, Liu H T, Lv H B, Gong T C, Long S B and Liu Q 2015 *IEEE International Electron Devices Meeting (IEDM)*, December 7–9, 2015, Washington, DC, USA, p. 10.2.1

[75] Liu W J, Tran X A, Yu H Y and Sun X W 2013 *ECS Solid State Lett.* **2** Q35

[76] Kim K H, Jo S H, Gaba S and Lu W 2010 *Appl. Phys. Lett.* **96** 053106

[77] Lv H B, Li Y T, Liu Q, Long S B, Li L and Liu M 2013 *IEEE Electron Device Lett.* **34** 229

[78] Subhechha S, Degraeve R, Roussel P, Goux L and Kar G S 2019 *IEEE International Memory Workshop (IMW)*. IEEE, May 12–15, 2019, Monterey, CA, USA, p. 1

[79] Hsu C W, Wan C C, Wang I T, Chen M C, Lo C L and Lee Y J 2013 *IEEE International Electron Devices Meeting*, December 9–11, 2013, Washington, DC, USA, p. 10.4.1

[80] Tran X, Zhu W, Liu W, Yeo Y, Nguyen B and Yu H 2013 *IEEE Transactions on Electron Devices* **60** 391

[81] Linn E, Rosezin R, Kügeler C and Waser R 2010 *Nat. Mater.* **9** 403

[82] Subhechha S, Degraeve R, Belmonte A, Goux L, Donadio G L, Roussel P, Meyer K D, Houdt J V and Kar G S 2018 *IEEE International Memory Workshop (IMW)* May 13–16 Kyoto, Japan, p. 1

[83] Schonhals A, Mohr J, Wouters D J, Waser R and Menzel S 2017 *IEEE Electron Device Lett.* **32** 449

[84] Banerjee W, Zhang X M, Luo Q, Lv H B, Liu Qi, Long S B and Liu M 2018 *Advanced Electronic Materials* **4** 1700561

[85] Jana D, Samanta S, Maikap S and Cheng H M 2016 *Appl. Phys. Lett.* **108** 011605

[86] Lim D H, Kim G Y, Song J H, Jeong K S, Ko D H and Cho M H 2015 *Scientific Reports* **5** 15374

[87] Yoon K J, Song S J, Seok J K, Yoon J H, Park T H and Kwon D E 2014 *Nanoscale* **6** 2161

[88] Govoreanu B, Clima S, Radu I P, Chen Y Y, Wouters D J and Jurczak M 2013 *IEEE Transactions on Electron Devices* **60** 2471

[89] Dai L, Wu H Q, Gao B and Qian H 2016 *International Symposium on VLSI Technology, Systems and Application (VLSI-TSA)*, Apirl 25–27, 2016, Hsinchu, Taiwan, p. 1

[90] Yu S, Guan X and Wong H S P 2012 *IEEE Transactions on Electron Devices* **59** 1183

[91] Wang Z W, Kang J, Fang Y C, Yu Z Z, Xue Y, Cai Y M, Wang Y Y and Huang R 2016 *16th Non-Volatile Memory Technology Symposium (NVMTS)*, October 17–19, 2016, Pittsburgh, PA, USA, p. 1

[92] Grossi A, Nowak E, Zambelli C, Pellissier C, Bernasconi S, Cibrario G, Hajjam K E, Crochemore R, Nodin J F, Olivo P and Perniola L 2016 *IEEE International Electron Devices Meeting (IEDM)*, December 3–7, 2016, San Francisco, CA, USA, p. 4.7.1

[93] Chen H Y, Yu S, Gao B, Huang P, Kang J and Wong H S P 2012 *IEEE International Electron Devices Meeting (IEDM)*, December 10–13, 2012, San Francisco, CA, USA, p. 20.7.1

[94] Yu S, Guan X and Wong H S P 2011 *IEEE International Electron Devices Meeting (IEDM)*, December 5–7,2011, Washington, DC, USA, p. 17.3.1

[95] Perez E, Grossi A, Zambelli C, Olivo P, Roelofs R and Wenger C 2016 *IEEE Electron Device Lett.* **38** 175

[96] Luo Q, Xu X X, Gong T C, Lv H B, Dong D N, Ma H L, Yuan P, Gao J F, Liu J, Yu Z A, Li J F, Long S B, Liu Q and Liu M 2017 *IEEE International Electron Devices Meeting (IEDM)*, December 2–6, 2017, San Francisco, CA, USA, p. 2.7.1

[97] Li C, Han L, Jiang H, Jang M H, Lin P, Wu Q, Barnell M, Yang J J, Xin H L and Xia Q F 2017 *Nat. Commun.* **8** 15666

[98] Jiang Z, Wang Z, Xin Z, Fong S, Qin S, Chen H Y, Ahn C Y, Cao J, Nishi Y and Wong H S Philip 2016 *IEEE International Electron Devices Meeting (IEDM)*, December 3–7, 2016, San Francisco, CA, USA, p. 21.3.1

[99] Jiang R, Han Z and Du X 2016 *Microelectronics Reliability* **63** 37

[100] Yu S, Gao B, Dai H, Sun B, Liu L F and Liu X 2010 *Electrochemical and Solid-State Letters* **13** H36

[101] Bin G, Bing C and Rui L and Zhang F F 2014 *IEEE Transactions on Electron Devices* **61** 1377

[102] Lee J, Shin J, Lee D, Lee W, Jung S and Jo M 2010 *IEEE International Electron Devices Meeting (IEDM)*, December 6–8, 2010, San Francisco, CA, USA, p. 19.5.1

[103] Chen H Y, Tian H, Gao B, Yu S M, Liang J and Kang J F 2012 *IEEE International Electron Devices Meeting (IEDM)*, December 10–13, 2012, San Francisco, CA, USA, p. 20.5.1

[104] Song Y, Magyari-Kope B, Lin Y Y and Nishi Y 2017 *IEEE International Memory Workshop* May 14–17, 2017, Monterey, CA, USA, p. 1

[105] Chang W Y, Huang H W, Wang W T, Hou C H, Chueh Y L and He J H 2012 *Journal of The Electrochemical Society* **159** G29

[106] Gao B, Zhang H, Yu S, Sun B, Liu L F, Liu X Y 2009 *Symposium on VLSI Technology* June 15–17, 2009, Kyoto, Japan, p. 30

[107] Zhang H, Liu L F, Gao B, Qiu Y, Liu X Y and Lu J 2011 *Appl. Phys. Lett.* **98** 042105

[108] Liu L F, Chen B, Gao B, Zhang F F, Chen Y S and Liu X Y 2011 *Appl. Phys. A* **102** 991

[109] Liu L F, Kang J F, Xu N, Sun X, Chen C, Sun B Wang Y, Liu X Y, Zhang X and Han R Q 2008 *Jpn. J. Appl. Phys.* **47** 2701

[110] Chen L, Gou H Y, Sun Q Q, Zhou P, Lu H L and Wang P F 2011 *IEEE Electron Device Lett.* **32** 794

[111] Chen G, Song C, Chen C, Gao S, Zeng F and Pan F 2012 *Adv. Mater.* **24** 3515

[112] Liu Q, Long S B, Wang W, Tanachutiwat S, Li Y T and Wang Q 2010 *IEEE Electron Device Lett.* **31** 1299

[113] Hsu C C, Lin Y S, Cheng C W and Jhang W C 2020 *IEEE Transactions on Electron Devices* **67** 3
[114] Liu H T, Lv H B, Yang B, Xu X X, Liu R Y and Liu Q 2014 *IEEE Electron Device Lett.* **35** 1224
[115] Kim K M, Lee S R, Kim S, Chang M and Hwang C S 2015 *Advanced Functional Materials* **25** 1527
[116] Fantini A, Goux L, Degraeve R, Wouters D, Raghavan N and Kar G 2013 *IEEE International Memory Workshop (IMW)*, May 26–29, 2013, Monterey, CA, USA, p. 30
[117] Bai Y, Wu H, Wu R, Zhang Y, Deng N, Yu Z and Qian H 2014 *Scientific Reports* **4** 5780
[118] Hsu C W, Wang Y F, Wan C C, Wang I T, Chou C T and Lai W L 2014 *Nanotechnology* **25** 165202
[119] Raghavan N 2014 *Microelectronics Reliability* **54** 2253
[120] Wiefels S, Bottger U, Menzel S, Wouters D J and Waser R 2020 *IEEE International Memory Workshop (IMW)*, 17–20 May 2020, Dresden, Germany, p. 1
[121] Raghavan N, Pey K L, Frey D D and Bosman M 2014 *IEEE International Reliability Physics Symposium*, June 1–5, 2014, Waikoloa, HI, USA, p. MY.9.1
[122] Lu Y, Chen B, Gao B, Fang Z, Fu Y H and Yang J Q 2012 *IEEE International Reliability Physics Symposium (IRPS)*, Apirl 15–19, 2012, Anaheim, CA, USA, p. MY.4.1–MY.4.4
[123] Lv H B, Xu X X, Liu H T, Liu R Y, Liu Q, Banerjee W, Sun H, Long S, Li L and Liu M and Banerjee W 2015 *Scientific Reports* **5** 7764
[124] Lin Y H, Ho Y H, Lee M H, Wang C H, Lin Y Y, Lee F M, Hsu K C, Tseng P H, Lee D Y, Chiang K H, Wang K C, Tseng T Y and Lu C Y 2017 *IEEE International Electron Devices Meeting (IEDM)*, December 2–6, 2017, San Francisco, CA, USA, p. 2.5.1
[125] Chen Y Y, Roelofs R, Redolfi A, Degraeve R, Crotti D and Fantini A 2014 *Symposium on VLSI Technology (VLSI-Technology): Digest of Technical Papers*, June 9–12, 2014, Honolulu, HI, USA, p. 1
[126] Lee H Y, Chen Y S, Chen P S, Gu P Y, Hsu Y Y and Wang S M 2010 *IEEE International Electron Devices Meeting (IEDM)*, December 6–8, 2010, San Francisco, CA, USA, p. 19.7.1
[127] Chen B, Lu, Gao B, Fu Y H, Zhang F F and Huang P 2011 *International Electron Devices Meeting Technical Digest*, December 5–7, 2011, Washington, DC, USA, p. 12.3.1
[128] Chang Y F, O'Donnell J A, Acosta T, Kotlyar R, Chen A, Quintero P A, Strutt N, Golonzka O, Connor C and Hicks J 2020 *IEEE International Reliability Physics Symposium (IRPS)*, 28 April–30 May 2020, Dallas, TX, USA, USA, p. 1
[129] Gao B, Kang J F, Zhang H W, Sun B, Chen B and Liu L F 2010 *Proceedings of the European Solid State Device Research Conference*, September 14–16, 2010, Sevilla, Spain, p. 392
[130] Gao B, Zhang H W, Chen B, Liu L F, Liu X Y and Han R Q 2011 *IEEE Electron Device Lett.* **32** 276
[131] Wei Z, Takagi T, Kanzawa Y, Katoh Y, Ninomiya T and Kawai K 2012 *4th IEEE International Memory Workshop*, May 20–23, 2012, Milan, Italy, p. 1
[132] Chen C Y, Fantini A, Degraeve R, Redolfi A, Groeseneken G, Goux L and Kar G S 2016 *IEEE International Electron Devices Meeting (IEDM)*, December 3–7, 2016, San Francisco, CA, USA, p. 4.6.1
[133] Zhao Y Q, Xu J, Yang J, Xue X Y, Lin Y Y and Sim J 2016 *in IEEE International Symposium on Circuits & Systems*, March 22–25, 2016, Montreal, QC, Canada, p. 2807
[134] Wang W, Yang H, Zhuo V Y, Li M, Chua E K and Yu J 2016 *in International Symposium on VLSI Technology, Systems and Application*, Apirl 25–27, 2016, Hsinchu, Taiwan, p. 1–2
[135] Xu X X, Lv H B, Liu H T, Gong T C, Wang G M and Zhang M Y 2015 *IEEE Electron Device Lett.* **36** 129
[136] Cheng C H, Chin A and Yeh F S 2010 *Symposium on VLSI Technology*, June 15–17, 2010, Honolulu, HI, USA, p. 85
[137] Yu S, Chen H Y, Deng Y, Gao B, Jiang Z and Kang J 2013 *Symposium on VLSI Technology (VLSIT)*, June 11–13, 2013, Kyoto, Japan, p. T158
[138] Sakamoto T, Tada M, Banno N, Tsuji Y, Saitoh Y and Yabe Y 2009 *Symposium on VLSI Technology*, June 15–17, 2009, Kyoto, Japan, p. 130
[139] Joo W, Lee J H, Choi S M, Kim H D and Kim S 2016 *Journal of Nanoscience and Nanotechnology* **16** 11391
[140] Lee S R, Kim Y B, Chang M, Kim K M, Lee C B and Hur J H 2012 *Symposium on VLSI Technology (VLSIT)*, June 12–14, 2012, Honolulu, HI, USA, p. 71
[141] Tsai C L, Xiong F, Pop E and Shim M 2013 *ACS Nano* **7** 5360
[142] Li H, Chen H Y, Chen Z, Chen B, Liu R and Qiu G 2014 *IEEE International Reliability Physics Symposium*, June 1–5, 2014,Waikoloa, HI, USA, p. MY.3.1
[143] Zhirnov V V, Meade R, Cavin R K and Sandhu G 2011 *Nanotechnology* **22** 254027
[144] Govoreanu B, Kar G, Chen Y, Paraschiv V, Kubicek S and Fantini A 2011 *IEEE International Electron Devices Meeting (IEDM)*, December 5–7, 2011, Washington, DC, USA, p. 31.6.1
[145] Lee M J, Han S, Jeon S H, Park B H, Kang B S and Ahn S E 2009 *Nano Lett.* **9** 1476
[146] Woźniak S, Pantazi A, Sidler S, Papandreou N, Leblebici Y and Eleftheriou E 2017 *IEEE Transactions on Circuits and Systems II: Express Briefs* **64** 1342
[147] Mead C 1990 *Proc. IEEE* **78** 1629
[148] Suri M, Bichler O, Querlioz D, Cueto O, Perniola L and Sousa V 2011 *IEEE International Electron Devices Meeting (IEDM)*, December 5–7, 2011, Washington, DC, USA, p. 4.4.1
[149] Pickett M D, Medeiros-Ribeiro G and Williams R S 2013 *Nat. Mater.* **12** 114
[150] Woo J, Moon K, Song J, Kwak M, Park J and Hwang H 2016 *IEEE Transactions on Electron Devices* **63** 5064
[151] Yu S M, Gao B, Fang Z, Yu H, Kang J and Wong H S P 2013 *Adv. Mater.* **25** 1774
[152] Kuzum D, Yu S and Wong H S P 2013 *Nanotechnology* **24** 382001
[153] Li Y, Zhong Y P, Zhang J J, Xu L, Wang Q, Sun H J, Tong H, Cheng X M and Miao X S 2014 *Scientific Reports* **4** 4906
[154] Chen L, Wang T Y, Dai Y W, Cha M Y, Zhu H, Sun Q Q, Ding S J, Zhou P, Chua L and Zhang D W 2018 *Nanoscale* **10** 15826
[155] Oh S, Kim T, Kwak M, Song J, Woo J, Jeon S, Yoo I K and Hwang H S 2017 *IEEE Electron Device Lett.* **38** 732
[156] Zhao W, Agnus G, Derycke V, Filoramo A, Bourgoin J and Gamrat C 2010 *Nanotechnology* **21** 175202
[157] Tanaka H, Akai-Kasaya M, TermehYousefi A, Hong L, Fu L X, Tamukoh H, Tanaka D, Asai T and Ogawa T 2018 *Nat. Commun.* **9** 1
[158] Kim K, Chen C L, Truong Q, Shen A M and Chen Y 2013 *Adv. Mater.* **25** 1693
[159] Chang T, Jo S H and Lu W 2011 *ACS Nano* **5** 7669
[160] Ohno T, Hasegawa T, Tsuruoka T, Terabe K, Gimzewski J K and Aono M 2011 *Nat. Mater.* **10** 591
[161] Woo J, Moon K, Song J, Lee S, Kwak M, Park J, Hwang H 2016 *IEEE Electron Device Lett.* **37** 994
[162] Suri M, Bichler O, Querlioz D, Palma G, Vianello E and Vuillaume A 2012 *IEEE International Electron Devices Meeting (IEDM)*, December 10–13, 2012, San Francisco, CA, USA, p. 10.3.1
[163] Tosson A M, Yu S, Anis M H and Wei L 2017 *IEEE Computer Society Annual Symposium on VLSI (ISVLSI)*, July 3–5, 2017, Bochum, Germany, p. 62
[164] Lee W, Park J, Kim S, Woo J, Shin J and Lee D 2012 *Appl. Phys. Lett.* **100** 142106
[165] Abbott L F and Nelson S B 2000 *Nature Neuroscience* **3** 1178
[166] Ambrogio S, Balatti S, Milo V, Carboni R, Wang Z Q, Calderon A, Ramaswamy N and Ielmini D 2016 *IEEE Transactions on Electron Devices* **63** 1508
[167] Padovani A, Woo J, Hwang H and Larcher L 2018 *IEEE Electron Device Lett.* **39** 672
[168] Dan Y and Poo M M 2004 *Neuron* **44** 23
[169] Bi G Q, Poo M M 1998 *The Journal of Neuroscience* **18** 10464
[170] Jeong D S, Kim I, Ziegler M and Kohlstedt H 2013 *RSC Advances* **3** 3169
[171] Yan X B, Zhao J H, Liu S, Zhou Z Y, Qi L, Chen J S and Liu X Y 2018 *Advanced Functional Materials* **28** 1705320
[172] Yang X, Cai Y, Zhang Z X, Yu M X and Huang R 2015 *15th Non-Volatile Memory Technology Symposium (NVMTS)*, October 12–14, 2015, Beijng, China, p. 1
[173] Sung C, Lim S, Kim H, Kim T, Moon K, Song J, Kim J J and Hwang H 2018 *Nanotechnology* **29** 115203
[174] Zeng F, Li S Z, Yang J, Pan F and Guo D 2014 *RSC Advances* **4** 14822
[175] Rubin D C, Hinton S and Wenzel A 1999 *Journal of Experimental Psychology: Learning, Memory, and Cognition* **25** 1161

[176] Park S, Sheri A, Kim J, Noh J, Jang J and Jeon M 2013 *IEEE International Electron Devices Meeting (IEDM) Technical Digest*, December 9–11, 2013, Washington, DC, USA, p. 25.6.1

[177] Lim S, Kwak M and Hwang H 2019 *Nanotechnology* **30** 455201

[178] Sakellaropoulos D, Bousoulas P, Nikas G, Arvanitis C and Tsoukalas D 2020 *Microelectronic Engineering* **229** 111358

[179] Ryu H and Kim S 2020 *Nanomaterials* **10** 1550

[180] Pan W Q, Chen J, Kuang R, Li Y, He Y H, Feng G R, Duan N, Chang T C and Miao X S 2020 *IEEE Transactions on Electron Devices* **67** 895

[181] Yu S M, Li Z W, Chen P Y, Wu H Q, Gao B, Wang D L, Wu W and Qian H 2016 *IEEE International Electron Devices Meeting (IEDM)*, December 3–7, 2016, San Francisco, CA, USA, p. 16.2.1

[182] Zamarreño-Ramos C, Camuñas-Mesa L A, Pérez-Carrasco J A, Masquelier T, Serrano-Gotarredona T and Linares-Barranco B 2011 *Frontiers in Neuroscience* **5** 26

[183] Bayat F M, Prezioso M, Chakrabarti B Nili H, Kataeva I and Strukov D 2018 *Nat. Commun.* **9** 2331

[184] Yao P, Wu H Q, Gao B, Tang J S, Zhang Q T, Zhang W Q, Yang J J and Qian H 2020 *Nature* **577** 641

[185] Wang I T, Lin Y C, Wang Y F, Hsu C W and Hou T H 2014 *IEEE International Electron Devices Meeting (IEDM)*, December 15–17, 2014, San Francisco, CA, USA, p. 28.5.1

[186] Wang I T, Chang C C, Chiu L W, Chou T and Hou T H 2016 *Nanotechnology* **27** 36

[187] Wang T Y, Meng J L, Chen L, Zhu H, Sun Q Q, Ding S J, Bao W Z and Zhang D W 2020 *InfoMat* **1**

[188] Li Z W, Chen P Y, Xu H and Yu S M 2017 *IEEE Transactions on Electron Devices* **64** 6

[189] Li H T, Li K S, Lin C H, Hsu J L, Chiu W C, Chen M C, Wu T T, Sohn J, Eryilmaz S B, Shieh J M, Yeh W K and Philip Wong H S 2016 *IEEE Symposium on VLSI Technology*, June 14–16, 2016, Honolulu, HI, USA, 2016, p. 1

[190] An H Y, Ehsan M A, Zhou Z and Yi Y 2017 *International Symposium on Quality Electronic Design*, March 14–15, 2017, Santa Clara, CA, USA, 2017, p. 1

[191] Kim B and Li H 2020 *IEEE Computer Society Annual Symposium on VLSI (ISVLSI)*, July 6–8, 2020, Limassol, Cyprus, Cyprus, 2020, p. 258

[192] Prezioso M, Merrikh-Bayat F, Hoskins B, Adam G, Likharev K K and Strukov D B 2015 *Nature* **521** 61

[193] Likharev K K and Strukov D B 2005 *Introducing Molecular Electronics* (Springer) p. 447

[194] DeHon A and Likharev K K 2005 *Proceedings of the IEEE/ACM International conference on Computer-aided design*, November 6–10, 2005, San Jose, CA, USA, p. 375

[195] Likharev K K 2007 *Journal of Vacuum Science & Technology B* **25** 2531

[196] Borghetti J, Li Z, Straznicky J, Li X, Ohlberg D A and Wu W 2009 *Proc. Natl. Acad. Sci. USA* **106** 1699

[197] Tu D, Liu M, Wang W and Haruehanroengra S 2007 *Micro & Nano Lett.* **2** 40

[198] Strukov D B and Likharev K K 2005 *Nanotechnology* **16** 137

[199] Strukov D B and Likharev K K 2006 *Proceedings of the ACM/SIGDA 14th International Symposium on Field Programmable Gate Arrays*, p. 131–140

[200] Zhong Z, Wang D, Cui Y, Bockrath M W and Lieber C M 2003 *Science* **302** 1377

[201] Chen Y, Jung G Y, Ohlberg D A, Li X, Stewart D R and Jeppesen J O 2003 *Nanotechnology* **14** 462

[202] Choudhury A and Sikdar B K 2017 *in International Symposium on VLSI Design and Test* June 29–July 2, 2017, Roorkee, India, p. 217

[203] Strukov D B and Likharev K K 2007 *Journal of Nanoscience and Nanotechnology* **7** 151

[204] Xia Q, Robinett W, Cumbie M W, Banerjee N, Cardinali T J and Yang J J 2009 *Nano Lett.* **9** 3640

[205] Madhavan A, Sherwood T and Strukov D B 2018 *IEEE Transactions on Very Large Scale Integration (VLSI) Systems* **26** 2759

[206] Cheng K T T and Strukov D B 2012 *Proceedings of the ACM International Symposium on International Symposium on Physical Design*, p. 33

[207] Prezioso M, Zhong Y, Gavrilov D, Merrikh-Bayat F, Hoskins B, Adam G, Likharev K and Strukov D 2016 *IEEE International Symposium on Circuits and Systems (ISCAS)*, May 22–25, 2016, Montreal, QC, Canada, p. 177

[208] Likharev K K 2009 *AAAI Fall Symposium Series* p. 90