HfOx Based Vertical Resistive Random Access Memory for Cost-Effective 3D Cross-Point Architecture without Cell Selector

Hong-Yu Chen[1], Shimeng Yu[1], Bin Gao[1,2], Peng Huang[2], Jinfeng Kang[2#], and H.-S. Philip Wong[1*]
[1] Department of Electrical Engineering and Center for Integrated Systems, Stanford University, Stanford, CA 94305, USA
[2] Institute of Microelectronics, Peking University, Beijing 100871, China
Email: #kangjf@pku.edu.cn, *hspwong@stanford.edu

Abstract - Double-layer stacked HfOx vertical RRAM is demonstrated for 3D cross-point architecture using a cost-effective fabrication process. Electrode/oxide interface engineering using TiON layer results in non-linear I-V suitable for the selector-less array. The fabricated HfOx vertical RRAM shows excellent performances such as reset current (<50μA), switching speed (~50ns), switching endurance (>10^8 cycles), half-selected read disturbance immunity (>10^9 cycles), retention (>10^5s @125°C). Moreover, a unique write/read scheme is proposed for 3D cross-point architecture. Analysis shows that for such 3D selector-less array, a large R_{on} (~100kΩ) from the non-linear I-V helps reduce the sneak path current, and a low interconnect resistance using metal planes as word lines reduces the undesirable voltage drop on the interconnect. As a conservative estimate, simulation shows that Mb-scale array without cell selector is achievable.

I. Introduction

Metal oxide RRAM is a very promising candidate for future non-volatile memory application [1-2]. To compete with the ultra-high density 3D NAND FLASH [3-4], a technology path toward 3D stackable RRAM is needed. Instead of simply stacking horizontal RRAM, a vertical RRAM structure [5-6] is more attractive for 3D cross-point architecture in terms of the bit-cost. However, due to the limited space in 3D cross-point architecture, building an additional selector for each RRAM cell is challenging and it reduces device density. Unlike NAND FLASH, the random access of individual RRAM cells also imposes a unique challenge for 3D cross-point architecture design. To address these issues, we fabricated double-layer stacked HfOx vertical RRAM with relatively large R_{on} (~100kΩ) to eliminate the need for cell selector. Combined with a vertical transistor [7] as the bit-line selector, the random access challenge can be met.

II. Vertical RRAM Device Fabrication

Fig. 1 shows our cost-effective fabrication process: 1) Multiple stacked Pt(20nm)/SiO_2(30nm) are deposited by evaporation/LPCVD; 2) A trench (1~100 μm in size) is dry etched reaching through all the layers in one etch step down to the bottom SiO_2 layer; 3) 5nm HfOx is deposited by ALD conformally covering the sidewall of the trench; 4) 150nm TiN (with oxygen) is deposited by reactive sputtering to fill the trench as the pillar electrode; 5) the planar electrode Pt is exposed by dry etching. Two types of samples are fabricated: single-layer sample (1L) with one cell on the sidewall per trench; double-layer sample (2L) with two cells on the sidewall per trench. Electrical measurements are performed using Keithley 4200 and Agilent 81150A. Fig. 2 (a)-(e) show the TEM image of 1L cell, top cell and bottom cell in 2L. The trench is not perfectly vertical due to the limited etching capability in our university fab and is not a fundamental issue of the device structure. Fig. 2 (f)-(g) show the vertical RRAM composition profile through the sidewall by EDX analysis. High resolution EELS analysis is performed at the interface between TiN and HfOx (Fig. 2 (h)): oxygen bump shows up inside the TiN electrode, suggesting the formation of an interfacial TiON layer, which may cause a tunneling barrier and a large R_{on} for the RRAM cell. It is speculated that the conductive filament tends to form at the corner of the planar electrode due to the enhanced electric field. This corner effect always exists even if the trench is etched ideally vertical as shown in the simulation in Fig. 3.

III. Vertical RRAM Device Performance

Fig. 4 shows the forming process for 20 cells with ~4V forming voltage. Fig. 5 shows the typical DC I-V curves of the 1L sample, and the top cell and bottom cell of the 2L sample. Fig. 6 shows their resistance distribution by continuous pulse cycling (R_{on}~100kΩ with >10× resistance window). Fig. 7 shows their reset current distribution (<50μA). Fig. 8 shows their switching voltage distribution by pulse IV amplitude sweep (~3V). These results show a consistent switching characteristic among 1L sample, top cell and bottom cell in 2L sample, thus suggesting the potential of stacking multi-layer vertical RRAM using this cost-effective fabrication process. It is also noted that the I-V curve in LRS is non-linear (Fig. 9). The LRS current is insensitive to the change of temperature (Fig. 10). Thus the conduction in LRS is suggested to be dominated by

tunneling through an interfacial barrier (as indicated by the EELS analysis in Fig. 2 (h)). As a result, R_{on} (read @ 0.1V) is effectively raised up to ~100kΩ (Fig. 6). Device-to-device variation is reasonably controlled as shown in the switching voltage distribution from 10 different cells (Fig 11). The applied pulse width exponentially depends on the switching voltage (Fig. 12) agreeing with the field-driven oxygen migration switching mechanism [8]. The reliability is also examined: $>10^8$ switching endurance cycles (Fig. 13), $>10^5$s @125°C retention (Fig. 14) are demonstrated. Most importantly, the device can maintain its state $>10^9$ cycles @1.5V disturbance test (Fig. 15) (note that the switching voltage ~ 3V), which enables the "V/2" write scheme for cross-point architecture.

IV. 3D Cross-Point Array Analysis

Fig. 16 shows the schematic of the proposed 3D cross-point array architecture with $4F^2/m$ cell size (m is the number of stacked layers). The vertical RRAM cells are formed at the intersection of each plane (WL) and each vertical pillar (just as we fabricated in this work). Unlike 3D NAND FLASH [3-4], 3D RRAM array should enable randomly accessed individual cells. Thus we propose using a vertical transistor to serve as bit-line (BL) selector, whose gate is controlled by the select-line (SL). During write, a specific cell is selected (Fig. 17): Vw is applied on the selected cell's WL, and Vw/2 is applied on all the unselected cells' WL to avoid unintentional writing. To select the pillar where the selected cell is located, the SL of that pillar is turned on and the corresponding BL is grounded. During read, a row of cells (on the same SL line) on one plane are read out simultaneously (Fig. 18): Vr is applied on the selected cells' WL, SL that controls the selected cells is turned on, and the data of a row of cells are read out by the sense amplifiers. In general, for the cross-point architecture without selector, the array size is mainly limited by the ratio between the interconnect resistance and the memory R_{on} [9]. A large R_{on} is beneficial for reducing the voltage drop on interconnect and mitigating the sneak path current. The resistor network is simulated in SPICE for our proposed 3D cross-point architecture with vertical RRAM at 20nm technology node. The simulation is done by simplifying the 3D array into a collection of 2D slices in x-z plane. The interconnect resistance between the neighbor cells in the vertical pillar is estimated to be 2.5Ω for 20nm node according to the ITRS 2011, and the interconnect resistance between the neighbor cells in the x-y plane is estimated to be 1Ω. This simplification is only a rough estimate because the interconnect resistance in the xy plane should scale up sub-linearly (due to the current path spreading) instead of linearly (as in the 2D resistor network) when the array size increases. The larger the R_{on} of the memory device, the larger the write/read margin (Fig. 19 & 20). For R_{on} values (~100kΩ) in this work, up to Mb-scale array size could be achieved without cell selector. Since we overestimate the plane interconnect resistance, the Mb-scale array size should be a conservative estimate. However, even with this overestimation, 3D cross-point architecture has larger write/read margins than the corresponding 2D cross-point architecture (Fig. 21 & 22) with the same number of memory bits due to a lower WL plane resistance (as opposed to WL wire resistance for a 2D cross-point array). Also increasing the number of stacked layers only slightly degrades the read/write margin (Fig. 23 & 24), and 64 layers could be stacked with reasonable read/write margin.

V. Conclusion

The achievements of this work include: 1) A cost-effective 3D RRAM fabrication process is demonstrated using double-layer vertically stacked HfOx-based RRAM; 2) TiON interfacial layer is introduced for obtaining non-linear I-V for a selector-less operation; 3) Excellent and consistent switching characteristics suggest the potential of stacking more layers; 4) A unique write/read scheme is proposed for cell random access in 3D cross-point architecture; 5) Simulation suggests that a large R_{on} and a low WL plane resistance can enable Mb-scale array without cell selector, and further increasing non-linearity in RRAM I-V is expected to enable even larger memory array.

Acknowledgement

We thank H. Tian, C.-Y. Chen, and S. Gupta for the discussions on device fabrication. This work is supported in part by the Stanford NMTRI and 973 Program (2011CBA00602). This work was done in part at the Stanford Nanofabrication Facility (SNF), a member of the NSF-supported NNIN. B. Gao is supported in part by the Stanford School of Engineering China Research Exchange Program. S. Yu is additionally supported by the Stanford Graduate Fellowship.

Reference

[1] H.-S. P. Wong, *et al.*, *Proc. IEEE*, 100, 1951-1970, 2012.

[2] B. Govoreanu, *et al.*, *Tech. Dig. IEDM* , 729-732, 2011.

[3] J. Jang, *et al.*, *Tech. Dig. Symp. VLSI Technol.,* 192-193, 2009.

[4] R. Katsumata, *et al.*, *Tech. Dig. Symp. VLSI Technol.,* 136-137, 2009.

[5] W.-C. Chien, *et al.*, *Tech. Dig. Symp. VLSI Technol.,* 153-154, 2012.

[6] I. G. Baek, *et al.*, *Tech. Dig.. IEDM* , 737-740, 2011.

[7] N. Singh, *et al.*, *IEEE Trans. Electron Devices,* 55, 3107-3118, 2008.

[8] S. Yu, *et al.*, *Appl. Phys. Lett.,* 98, 103514, 2011.

[9] J. Laing, *et al.*, *IEEE Trans. Electron Devices,* 57, 2531-2538, 2010.

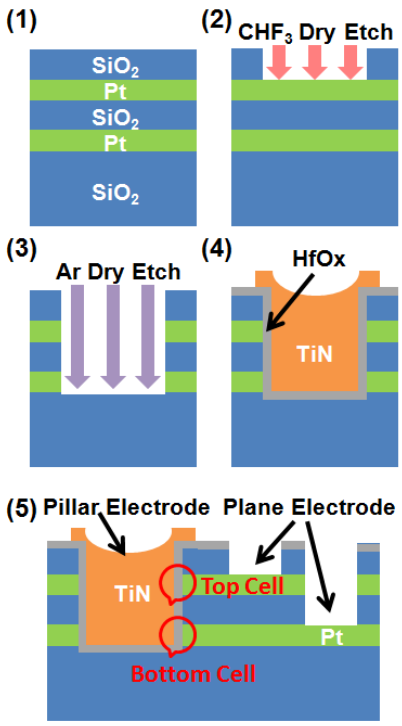

Fig. 1 The cost-effective process flow of 3D vertical RRAM: 1) Multiple Pt(20nm)/SiO_2(30nm) are deposited by evaporation/LPCVD; 2) A trench (1~100 μm in size) is dry etched down to the bottom SiO_2 layer; 3) 5nm HfOx is deposited by ALD conformally covering the sidewall of the trench; 4) 150nm TiN is deposited by sputtering to fill the trench as the pillar electrode; 5) the plane electrode Pt is exposed by dry etching.

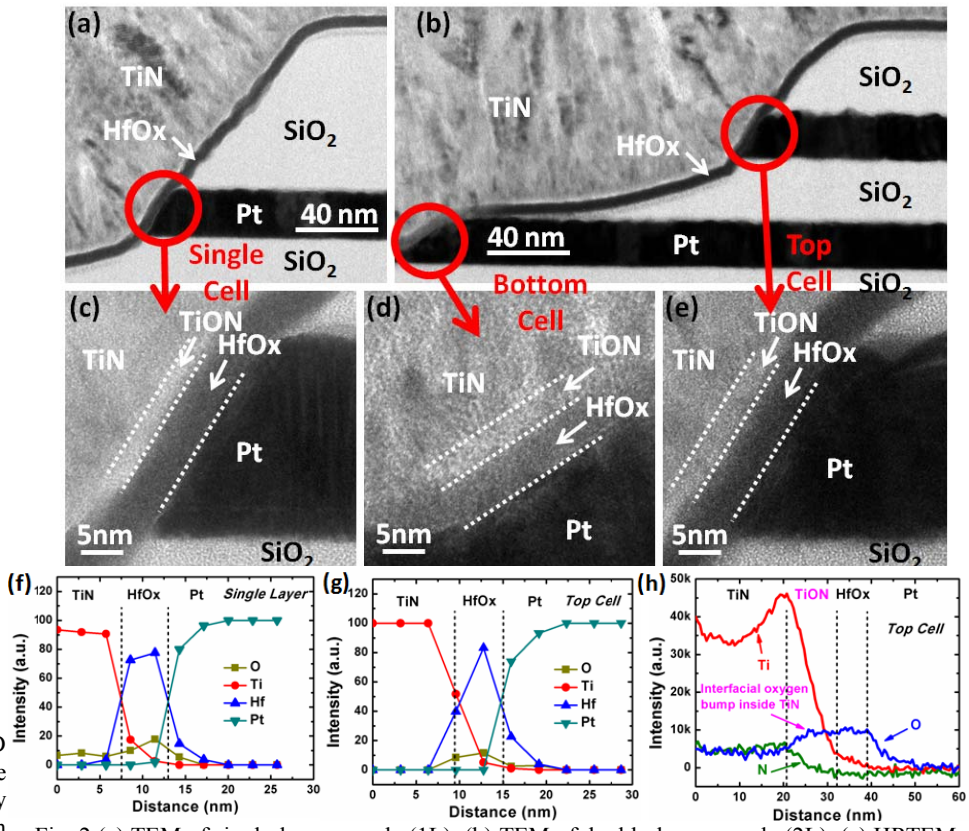

Fig. 2 (a) TEM of single-layer sample (1L); (b) TEM of double-layer sample (2L); (c) HRTEM of 1L cell; (d) HRTEM of bottom cell in 2L; (e) HRTEM of top cell in 2L; The trench is not perfectly sharp due to the limited etching capability in our university fab. (f) Energy dispersive X-ray spectroscopy (EDX) composition profile through the sidewall of 1L cell. (g) EDX composition profile through the sidewall of top cell in 2L. (h) Electron energy loss spectroscopy (EELS) composition profile through the sidewall of top cell in 2L. Oxygen bump shows up inside the TiN, suggesting the formation of an interfacial TiON layer, which may cause a tunneling barrier and a large Ron for the RRAM cell.

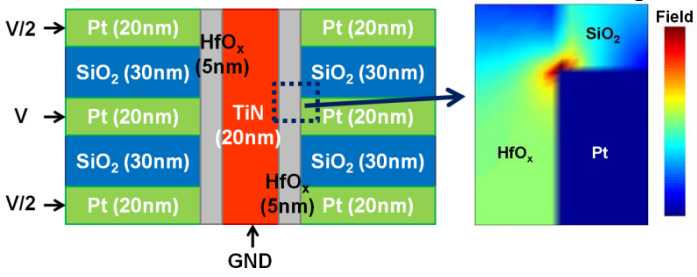

Fig. 3 Schematic of the vertical RRAM (left) and the simulated electric field distribution in the selected cell (right). During programming, TiN electrode is grounded and the voltage is applied on Pt electrode. The electric field in HfOx is strongest at the corner of plane electrode. Thus the conduction filament tends to form at the corner.

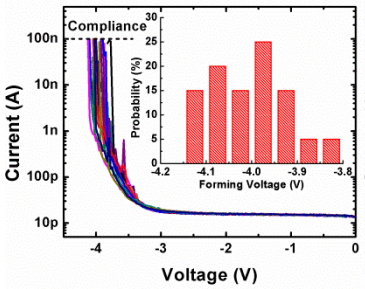

Fig. 4 Forming curve collected from 20 cells. 100nA compliance is applied. Inset: forming voltage distribution. Good uniformity of forming voltage is observed.

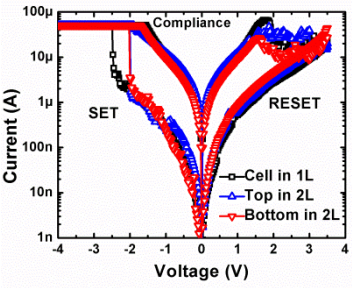

Fig. 5 Typical DC I-V switching characteristics. Consistent curves among the single-layer sample, (1L) top cell and bottom cell of the double-layer sample (2L).

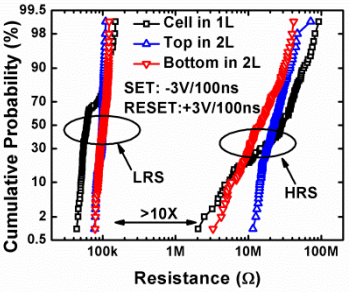

Fig. 6 Resistance distribution during continuous pulse cycling. LRS is ~100kΩ with a HRS/LRS window>10. Consistent resistance among single-layer cell (1L), top cell and bottom cell (2L).

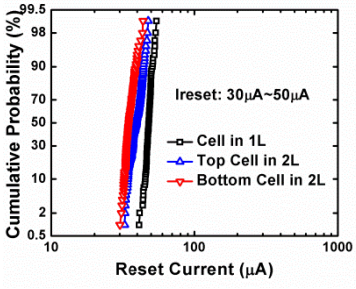

Fig. 7 Reset current distribution during continuous DC cycling. The reset current is 30μA~50μA. Consistent reset current among single-layer cell (1L), top cell and bottom cell (2L).

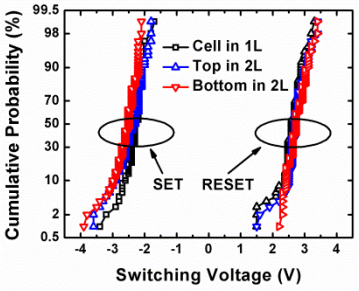

Fig. 8 Set and reset voltage distribution during continuous pulse I-V sweep cycling. Consistent switching voltage among single-layer cell (1L), top cell and bottom cell (2L).

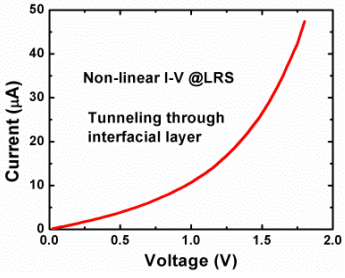

Fig. 9 I-V curve of LRS in linear scale. Non-linear I-V indicates an interfacial TiON tunneling layer exists (agreeing with the EELS analysis in Fig. 2 (h)).

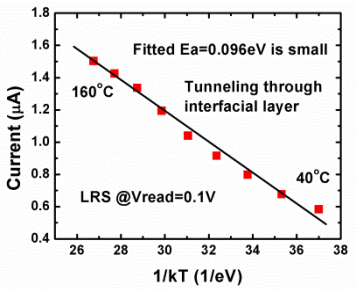

Fig. 10 Temperature dependence of the LRS resistance. A weak dependence (Ea~only 0.096V), also indicates tunneling current through an interfacial layer is the dominate mechanism.

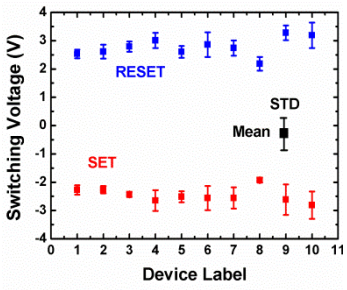

Fig. 11 Set and reset voltage distribution for 10 different cells. The mean switching voltage ~3V for different cells. The cycle to cycle (error bar) and device to device (x-axis) variation is reasonably controlled.

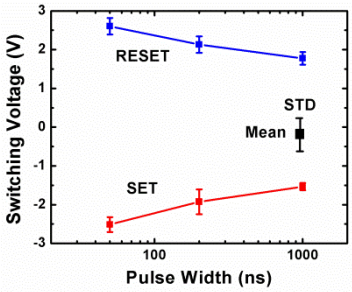

Fig. 12 Set and reset voltage as a function of pulse width. The pulse width required for programming exponentially decreases with increasing applied voltage. The device can switch using 50ns pulses with the voltage ~3V.

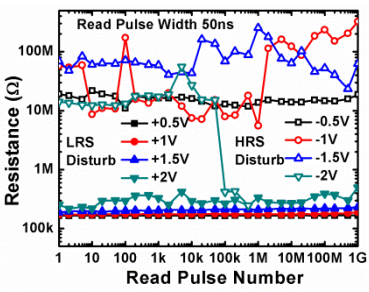

Fig. 13 Disturbance immunity test. The device can maintain its state under >10^9 pulse cycles with the disturb voltages up to ±1.5V. Such immunity enables the "V/2" write scheme for the cross-point array operation.

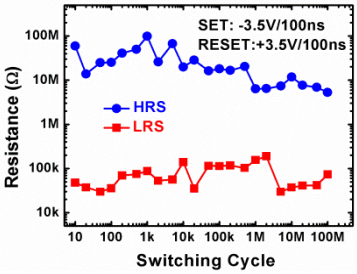

Fig. 14 Switching endurance test. The device can switch >10^8 pulse cycles.

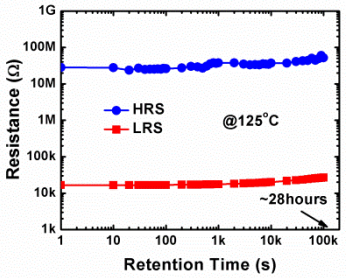

Fig. 15 Retention test. Both HRS and LRS can maintain their states >10^5 seconds (~28 hours) @ 125°C.

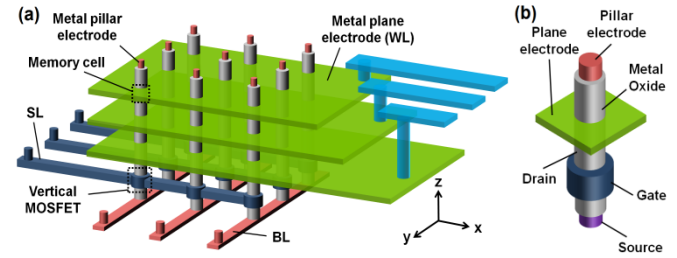

Fig. 16 Schematic view of our proposed 3D cross-point architecture (a) using the vertical RRAM cell demonstrated in the work and a vertical MOSFET transistor (b) as the bit-line selector to enable the random access capability of individual cells in the array.

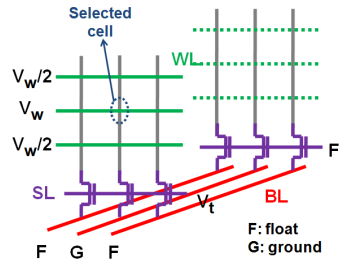

Fig. 17 Write scheme for the 3D cross-point architecture. Vw is applied on the selected cell's WL, and Vw/2 is applied on all the unselected cells' WL to avoid unintentional writing. SL of the selected cell's pillar is turned on and its BL is ground.

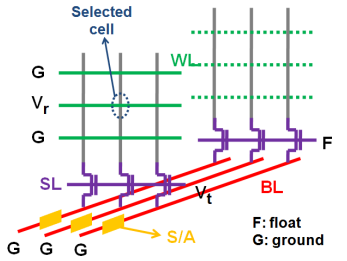

Fig. 18 Read scheme for the 3D cross-point architecture. Vr is applied on the selected cells' WL, SL that controls the selected cells is turned on, and the data of a row of cells are read out by the sense amplifier (S/A).

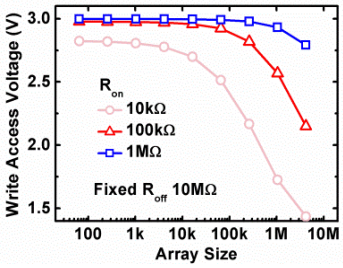

Fig. 19 Write margin for 3D cross-point architecture (16 layers of n×n xy planes) with different R_{on} and fixed R_{off} (~10MΩ). Larger R_{on} results in a smaller voltage drop on the wire resistance, thus retaining a larger write margin. For R_{on}~100kΩ in this work, Mb-scale size is achievable without cell selector.

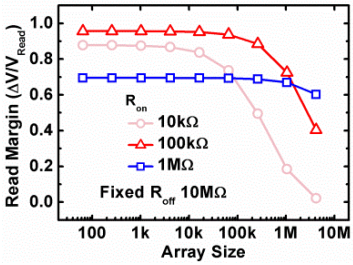

Fig. 20 Read margin for 3D cross-point architecture (16 layers of n× n xy planes) with different R_{on} and fixed R_{off} (~10MΩ). There is a trade-off: small R_{on} increases the sneak path current, while large R_{on} decreases the resistance window between HRS and LRS.

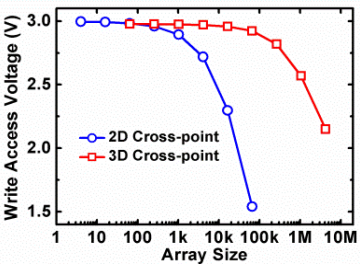

Fig. 21 Write margin comparison between 2D and 3D cross-point architecture (16 layers of n×n xy planes) with fixed R_{on} (~100kΩ) and R_{off} (~10MΩ). Much larger array size is achievable in 3D because of the reduced WL resistance as a plane.

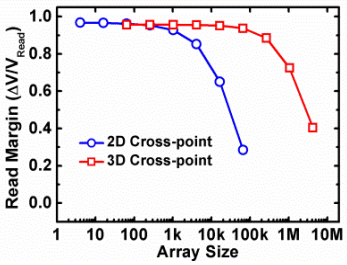

Fig. 22 Read margin comparison between 2D and 3D cross-point architecture (16 layers of n×n xy planes) with fixed R_{on} (~100kΩ) and R_{off} (~10MΩ). Much larger array size is achievable in 3D because of the reduced WL resistance as a plane.

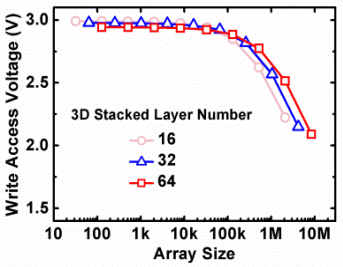

Fig. 23 Write margin for 3D cross-point architecture for different number of stacked layers (n×n xy planes) with fixed R_{on} (~100kΩ) and R_{off} (~10MΩ). Adding more layers has only a small penalty of the write margin thanks to the low WL plane resistance.

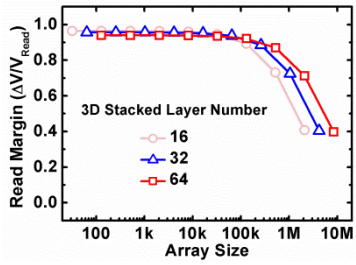

Fig. 24 Read margin for 3D cross-point architecture for different number of stacked layers (n×n xy planes) with fixed R_{on} (~100kΩ) and R_{off} (~10MΩ). Adding more layers has only a small penalty of the read margin thanks to the low WL plane resistance.