

Compact One-Transistor-N-RRAM Array Architecture for Advanced CMOS Technology

Chih-Wei Stanley Yeh, *Member, IEEE*, and S. Simon Wong, *Fellow, IEEE*

Abstract—For RRAM to be a cost-competitive candidate for high-density and high-capacity commercial products, some architectural-level challenges must be tackled. In this paper, research results that advance the design of high-density RRAM arrays are presented. We first focus on the scaling effects of on-chip interconnects on RRAM array performance. Due to the continuously shrinking process feature size, the voltage drop along the interconnect gradually reduces the voltage available to operate the RRAM device. To more efficiently analyze this effect for an arbitrary array size, a compact array model is developed. Simulations using this model determine the maximum achievable array size for future technology nodes. A compact, one-transistor-N-RRAM (1TNR) array architecture, with corresponding read/write and decoding schemes, that achieves high RRAM density is then introduced. A proof-of-concept 1T4R test chip with fully integrated RRAM devices is described. For this test chip, a particular sequence to form the cross-point RRAM array is presented. Measurement results of successful array operations demonstrate the feasibility and reliability of the proposed high-density architecture.

Index Terms—Array model, cross-point, flash, interconnect, multi-layer, nonvolatile memory, RRAM, 1TNR, 1T4R.

I. INTRODUCTION

FLASH memory chip has become ubiquitous. However, there is a concern about the future scalability of current floating gate flash technology. Many emerging technologies, such as PRAM (Phase-Change RAM), MRAM (Magnetoresistive RAM), FeRAM (Ferroelectric RAM), and RRAM (Resistive RAM), are being studied. Data collected in [1] shows that RRAM is particularly promising. Its read bandwidth is as good as NOR flash and its capacity is comparable to that of NAND flash. RRAM is composed of a programmable dielectric, typically metal-oxide, sandwiched between two metal electrodes. Initially, the dielectric is highly resistive ($G\Omega$), which is known as the high resistance state (HRS). When a sufficiently large voltage is applied across the dielectric, oxygen vacancies are generated in the dielectric, which supports the conduction of current in μA to mA range, and is known as the low resistance state (LRS). When a reverse voltage of sufficiently large magnitude is applied across the dielectric, the vacancies are back-filled and the RRAM returns to HRS. The transition from HRS to

LRS is known as SET, while the transition from LRS to HRS is known as RESET. In many RRAM technologies, a higher voltage is required to in the first SET cycle, which is known as FORMing.

In general, there are two ways to construct an RRAM array. One is the one-transistor-one-RRAM (1T1R) array, in which one dedicated transistor controls the access of one RRAM device, and the other is the cross-point array [2]–[4]. In 1T1R array, the cell size is determined by the transistor and is much larger than that of flash. Within the cross-point array RRAM cells are sandwiched between top bit lines (BLs) and word lines (WLs), without a dedicated transistor for each RRAM device. There is one access transistor connecting each BL (and WL) to the supporting circuits outside the array. For an array of M BLs and N WLs, the number of access transistors is only $M + N$. If all the access transistors could be physically fabricated underneath the RRAM devices, the unit RRAM cell size would be just $2F \times 2F = 4F^2$, the smallest possible.

While the cross-point array is very area-efficient, it is not easy to read or write individual RRAM devices effectively. The schematic on the left in Fig. 1 shows a general scheme to SET an RRAM device in a 1T1R array. The access transistors on the selected WL are turned on, the selected BL is applied with a voltage V_{SET} , and the selected source line (SL) is grounded. The SET current could be controlled by the optional current limiter I_{C1} or the gate voltage of the access transistor. The unselected BLs are grounded and hence there are no currents flowing through those unselected RRAM cells.

The schematic on the right in Fig. 1 presents a typical way to SET an RRAM device in a cross-point array:

- The selected bit line is applied with a voltage V_{SET} and the selected word line is grounded. That action imposes a V_{SET} across the selected cell (drawn as a black bold resistor).
- The unselected bit lines and word lines are all given a voltage $V_{SET}/2$. In this way the voltage across the unselected cells is either $V_{SET}/2$ or zero, preventing the false programming of the unselected cells.

The above *half V_{SET}* method prevents the false SETting of unselected RRAM cells. However, it is difficult to control the SET current in an RRAM cell in the cross-point array. As shown in the schematic, it is difficult to use a fixed current source to limit the current of the selected cell because the current source also needs to supply the current flowing into the other cells (the red bold resistors) on the same bit line. The currents flowing to the unselected cells are called the *leakage path currents*, and they are dependent not only on the number but also on the state of unselected cells.

Manuscript received August 24, 2014; revised November 19, 2014 and December 31, 2014; accepted January 26, 2015. Date of publication February 27, 2015; date of current version April 30, 2015. This paper was approved by Associate Editor Stefan Rusu.

The authors are with the Department of Electrical Engineering, Stanford University, Stanford, CA 94305 USA.

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/JSSC.2015.2402217

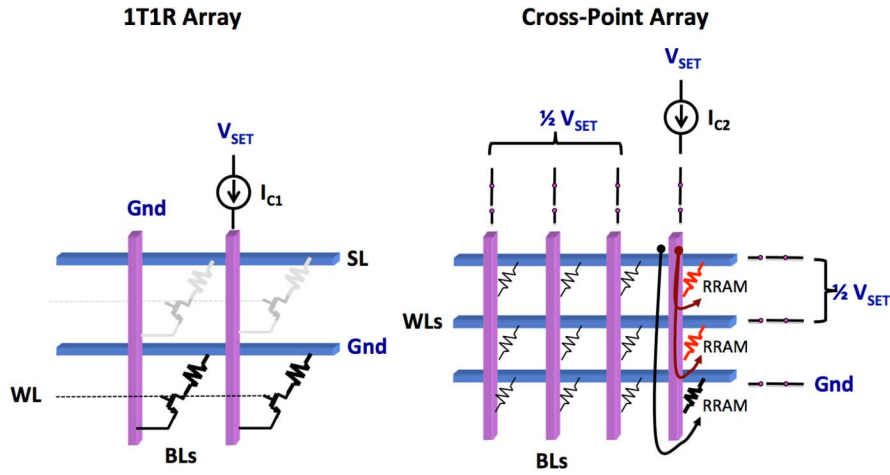
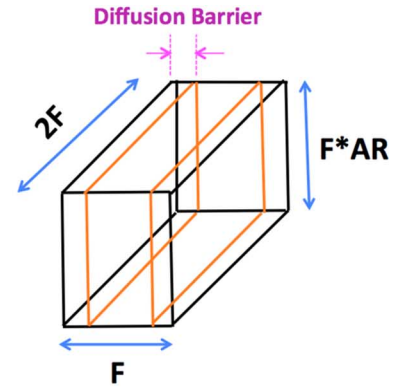


Fig. 1. SET scheme for 1T1R and cross-point arrays.

Pitch (nm)	17	19	21	24	27	30	34	38	43
Wire Width F (half pitch)(nm)	8.5	9.5	10.5	12.0	13.5	15.0	17.0	19.0	22.5
Diffusion Barrier (nm)	0.7	0.8	0.9	1.0	1.1	1.2	1.3	1.5	1.7
Aspect Ratio AR	2.1	2.1	2.1	2.1	2.0	2.0	2.0	2.0	2.0
Resistivity ($\mu\Omega\text{-cm}$)	14.0	13.0	12.0	11.0	10.0	9.0	8.0	8.0	7.0
Wire Resistance r_w (for $4F^2$ cell) (Ω)	19.7	16.5	13.8	11.0	8.8	7.1	5.6	5.0	3.9

(a)



(b)

Fig. 2. Interconnect in advanced technologies. (a) ITRS 2010 data for intermediate copper interconnect, (b) Interconnect cross-section.

To reduce the leakage path current, the array size, particularly the number of word lines, should be limited. Another means to reduce the leakage path current is to reduce the current of the unselected cells directly. This reduction can be accomplished either by reducing the voltage across the unselected cells to a level even less than $V_{SET}/2$ or by incorporating RRAM devices with nonlinear IV characteristics. Most of the RRAM devices published have a diode-like IV behavior, which reduces the leakage path currents in a cross-point array. To further reduce the leakage path current, some published works propose integrating diodes in series with the RRAM device, the *IDIR* cell. A current on/off ratio of more than 100 has been demonstrated [5], which could significantly reduce the leakage path current. However, bi-directional diodes are needed to work with bipolar RRAM devices. The diodes must sustain a high current density during programming, and will undoubtedly increase the overall operation voltage of the RRAM array.

Interconnect scaling effects is also a great concern in high-density and high-capacity RRAM. The decrease of the interconnect feature size not only introduces delay, but also limits the practical size of array. In this paper, the scaling effects on cross-point array performance will be discussed.

This paper consists of the following sections. Section II gives a detailed analysis of RRAM arrays in advanced technology,

emphasizing the effects of leakage path current and the interconnect wire resistance. Section III presents an one-transistor-N-RRAM (1TNR) array architecture that achieves high RRAM density comparable to cross-point array while minimizing the leakage paths at the same time. Read/write schemes for this new architecture are presented as well. Section IV validates the proposed 1TNR architecture by a proof-of-concept test chip. Finally, Section V summarizes this paper.

II. RRAM ARRAY IN ADVANCED TECHNOLOGY

Fig. 2(a) shows the data for intermediate copper interconnect, compiled from the ITRS 2010 roadmap [6]. The last row gives the resistance of a piece of interconnect with the shape illustrated in Fig. 2(b). The impact of increasing wire resistance with scaling as well as leakage path current on RRAM array is presented in this section.

A. RRAM Cell Resistance Model

Fig. 3 shows a resistance model of a cross-point RRAM cell that consists of an RRAM device, a piece of source line (SL), and a piece of bit line (BL).

- Assuming that the cell layout is an optimal $4F^2$ square, both the source line and the bit line are modeled as a re-

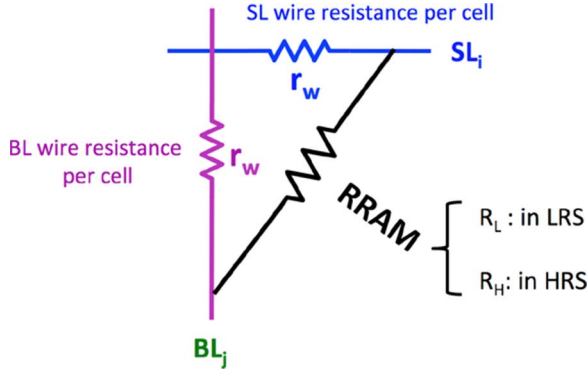


Fig. 3. RRAM cell interconnect resistance model.

sistor r_w whose value is shown in Fig. 2(a), depending on the interconnect feature size.

- The RRAM resistor shown in the model represents the RRAM device. It has two states:
 - 1) R_L : the RRAM device is in the low resistance state (LRS).
 - 2) R_H : the RRAM device is in the high resistance state (HRS).

B. RRAM Array Voltage Drop Analysis

One of the major effects of interconnect resistance upon the RRAM array operation is the excess voltage drop along the interconnect. These voltage drops reduce the available voltage across the RRAM device, rendering it difficult to be programmed. Between the SET and the RESET array operation, we will focus on RESET operation because the RESET current is higher than the SET current, thus inducing higher voltage drop along the interconnect.

Fig. 4(a) shows the $M \times N$ RRAM array to be analyzed in RESET operation. The horizontal lines are the source lines, the vertical lines are the bit lines, and there is an RRAM device at every cross point of a source line and a bit line. To carry out an interconnect voltage drop analysis we consider an extreme case:

- a RESET of every cell on SL1, where SL1 is the selected source line;
- all the cells on SL1 are in LRS.

This scenario introduces the largest voltage drop along the first source line and hence leaves the smallest voltage across the RRAM device at the top right corner, denoted as V_X . All the bit lines are fed with ground voltage while all the unselected source lines are applied with $V_{\text{RESET}}/2$. By substituting each RRAM cell by the cell model shown in Fig. 3, the full array model shown in Fig. 4(b) is obtained. In this array model, R_L and R_u represent the LRS RRAM on the first source line and the RRAM cells on the unselected source lines, respectively. In principle, based on the array model, a circuit schematic could be created and a SPICE (or equivalent) simulation could be performed to get the value of V_X for a given V_{RESET} , r_w , R_L , R_u , and array size.

Fig. 4(c) shows the compact model, derived from the complete model in Fig. 4(b) [7], to speed up the interconnect voltage drop analysis [8]. Having the compact model that works for arbitrary array size $M \gg 1$ and $N \gg 1$, closed-form equations

could be obtained to calculate V_X [7]. Fig. 5 shows the comparison between the V_X calculation results using the closed-form equations and the V_X circuit simulation results based on the complete array schematic. The results are similar for various array sizes and R_u/R_L ratios. This plot demonstrates the effectiveness of the proposed compact array model; it yields accurate results by simple calculation and works for a large range of array sizes. In Fig. 5 a green horizontal dashed line is drawn to indicate the level where $V_X/V_{\text{RESET}} = 0.5$. Where this line intersects the curves, three downward arrows are shown to point out the r_w/R_L value of each array size that generates $V_X = 0.5 \cdot V_{\text{RESET}}$. For instance, to get $V_X > 0.5 \cdot V_{\text{RESET}}$, the 32×32 array needs r_w/R_L to be less than $\approx 2 \times 10^{-3}$; for the 128×128 array r_w/R_L should be smaller than $\approx 10^{-4}$. Such a low r_w is impractical for the advanced technology nodes.

C. Solutions to Mitigate Voltage Drop Effects

To mitigate the interconnect voltage drop problem there are several solutions. The first one is to dynamically adjust the bias voltage, after checking whether the selected RRAM cells are programmed well. However, an impractically high voltage may be required to accommodate the interconnect voltage drop. A more effective solution is to physically partition a big array into multiple sub-arrays. For example, an $M \times M$ array could be divided into several $M \times N$ sub-arrays. The reduced number of bit-line could help to mitigate the voltage drop effect.

Fig. 6 shows the schematic and the compact model of a partitioned $M \times M$ array, in which only the last $M \times N$ sub-array is selected to be RESET. The array is in the worst case scenario in which the selected sub-array is the one farthest from the side where V_{RESET} is supplied. In this figure the R_u resistors are ignored for simplicity. This simplification is considered valid because it can be seen from the result in Fig. 5 that R_u/R_L ratio does not affect V_X value much.

Based on the model in Fig. 6 and the r_w values given in Fig. 2(a), for an 1024×1024 array partitioned into multiple sub-arrays, each of size $1024 \times N$, the plot in Fig. 7 can be generated. This plot has a black solid curve r_w that meets $V_X = 0.5 \cdot V_{\text{RESET}}$ for fixed $R_L = 100 \text{ k}\Omega$ and various N values. In addition, several horizontal dashed lines are drawn to point out the r_w values that correspond to the various interconnect feature sizes. At the intersections of the black solid line and the dashed lines, we can identify the N values that meet $V_X/V_{\text{RESET}} = 0.5$ for a given interconnect feature size F . These N values are summarized in Fig. 8, not only for $R_L = 100 \text{ k}\Omega$ but also for 10 k , 1 M and $10 \text{ M}\Omega$. This table shows that, for more advanced technology with small F , RRAM LRS resistance R_L should be increased to prevent too small an N value, which would make the circuits complex and hence not cost-effective. For instance, to have $N > 10$, R_L should be at least on the order of $100 \text{ k}\Omega$ for interconnect F size below 14 nm .

III. ONE-TRANSISTOR-N-RRAM ARCHITECTURE

The last section showed that a high R_L is necessary to construct a large cross-point sub-array. However, many RRAM devices published to date have R_L value of only around $10 \text{ k}\Omega$. Therefore, it is desired to have a cross-point array architecture that works well for small number of bit line. In this section, a

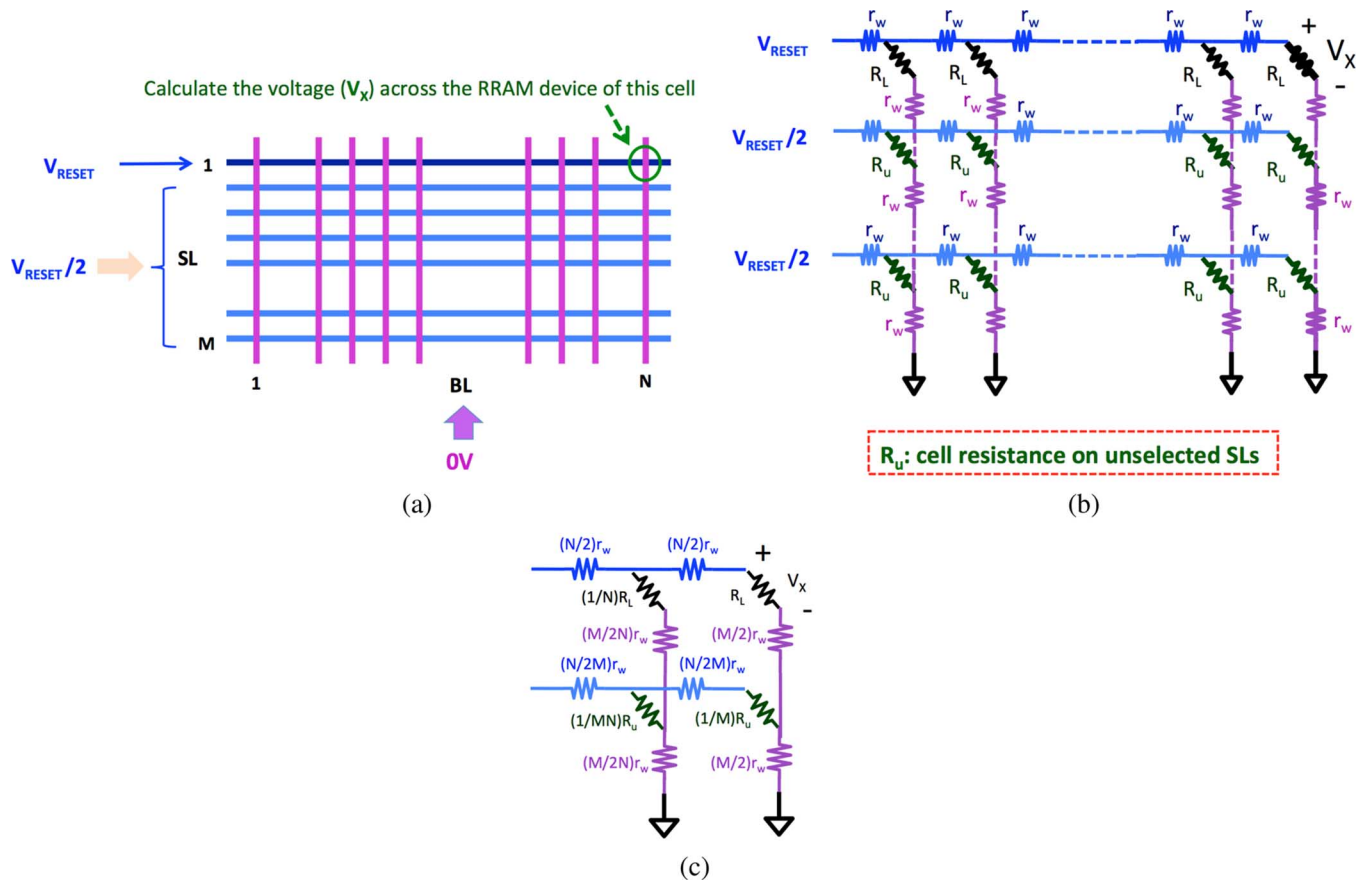


Fig. 4. Voltage drop analysis for an array in RESET operation. (a) Array schematic. (b) Array model for analysis. (c) Compact model for the $M \times N$ RRAM array.

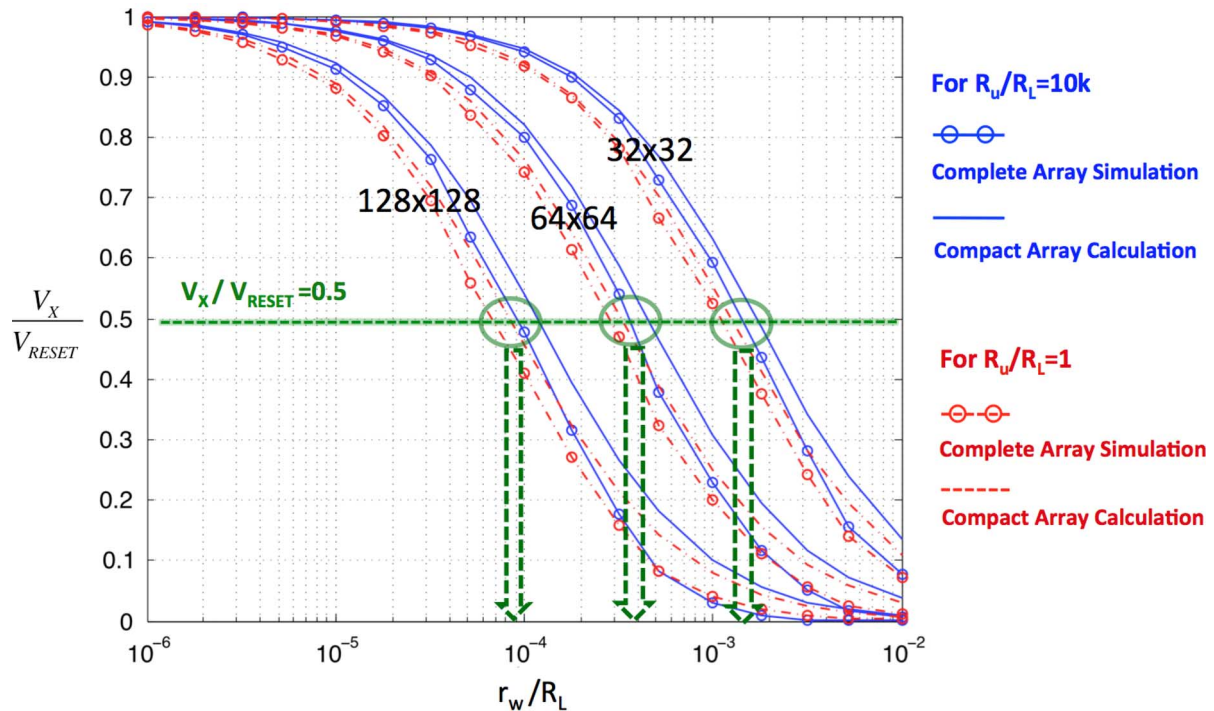


Fig. 5. Comparison between compact array model calculation and complete schematic simulation.

new one-transistor-four-RRAM (1T4R) array architecture with only four bit line is presented first to meet the need. A more

general one-transistor-N-RRAM (1TNr) scheme will then be discussed. Although 3D 1TNr architecture has been proposed,

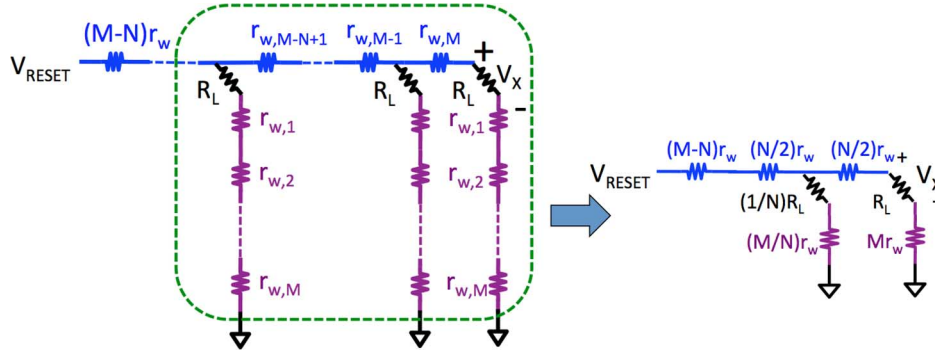


Fig. 6. Schematic and compact model for a partitioned $M \times M$ array. Only the last $M \times N$ array is activated.

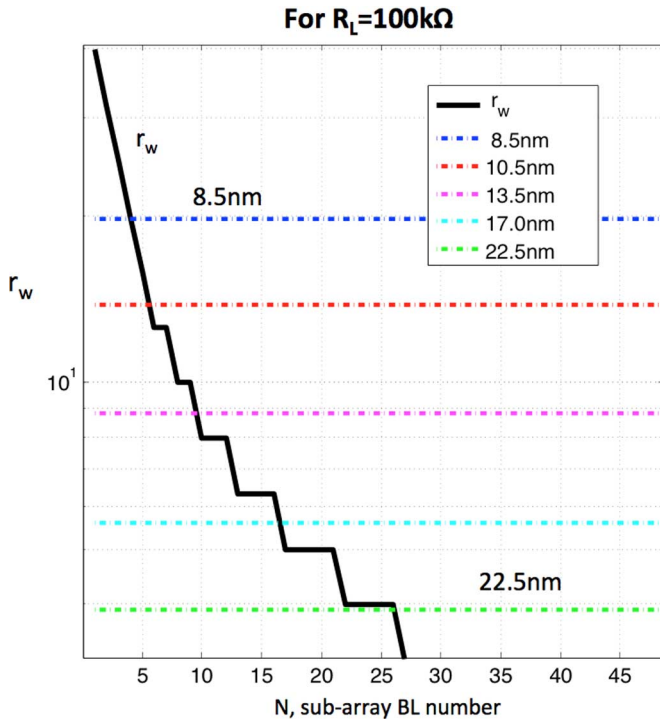


Fig. 7. Plot of N values to meet $V_X = 0.5 \times V_{\text{RESET}}$, for $R_L = 100 \text{ k}\Omega$.

N for different interconnect F and R_L

Interconnect Feature Size F (r_w)	8.5nm (19.7Ω)	10.5nm (13.8Ω)	13.5nm (8.8Ω)	17.0nm (5.6Ω)	22.5nm (3.9Ω)
$R_L=10\text{k}\Omega$	0	0	0	0	1
$R_L=100\text{k}\Omega$	4	6	10	17	26
$R_L=1\text{M}\Omega$	55	70	115	191	326
$R_L=10\text{M}\Omega$	1024	1024	1024	1024	1024

Fig. 8. Table of N to meet $V_X/V_{\text{RESET}} = 0.5$, for different F and R_L .

there has been no experimental demonstration of its feasibility [9]–[11]. In this paper, experimental results demonstrating 2D 1T4R architecture will be described.

A. 1T4R 4×4 Sub-Array

Fig. 9 shows the layout of a 4×4 sub-array and the corresponding schematic. All the critical feature sizes in the layout are the minimum values according to the MOSIS deep submicron lambda rule:

- metal width = metal spacing = $3\lambda = F$;
- poly gate width = 2λ ;
- contact size = 2λ ;
- active (diffusion) width = 3λ ;
- metal enclosure contact = 1λ ;
- contact spacing to poly gate = 2λ .

The 4×4 RRAM sub-array shown in Fig. 9, consisting of 16 RRAM cells sandwiched between four top bit lines (BL) and four word lines (WL), is fabricated on top of four select transistors whose gates are called the word selection lines (WSL). The bit lines run vertically, the word lines run horizontally, and there is one RRAM device at the cross-point of each bit line and each word line. The four RRAM cells in each row share a common bottom electrode that is connected, through vias, to the drain node of a minimally sized select transistor (1T4R). The top electrodes of these four cells are connected to four separate bit lines. Two adjacent select transistors that connect to two adjacent rows of RRAM cells share a common horizontal global word line (GWL). This sub-array is repeated horizontally and vertically to form a large RRAM array, as illustrated in Fig. 10.

In this architecture the select transistor has the minimum channel width and length so that the four transistors can fit into the area of the 4×4 sub-array. Hence, the effective RRAM cell size is $4F^2$. Because there are only three unselected RRAM devices along with one selected RRAM in the same row, independent of the size of cross-point array, the horizontal leakage path current is controlled.

B. Array Operation Schemes

The bit line that connects to the selected RRAM cell is called the selected bit line, the transistor that accesses the selected cell is called the selected transistor, and the global word line connecting to the selected transistor is called the selected global word line. All other bit lines, transistors, and global word lines are unselected. The RRAM cells are assumed to be bipolar switching devices.

Fig. 11 shows the scheme to SET a selected RRAM cell in the array. The gate of the selected transistor is first turned on

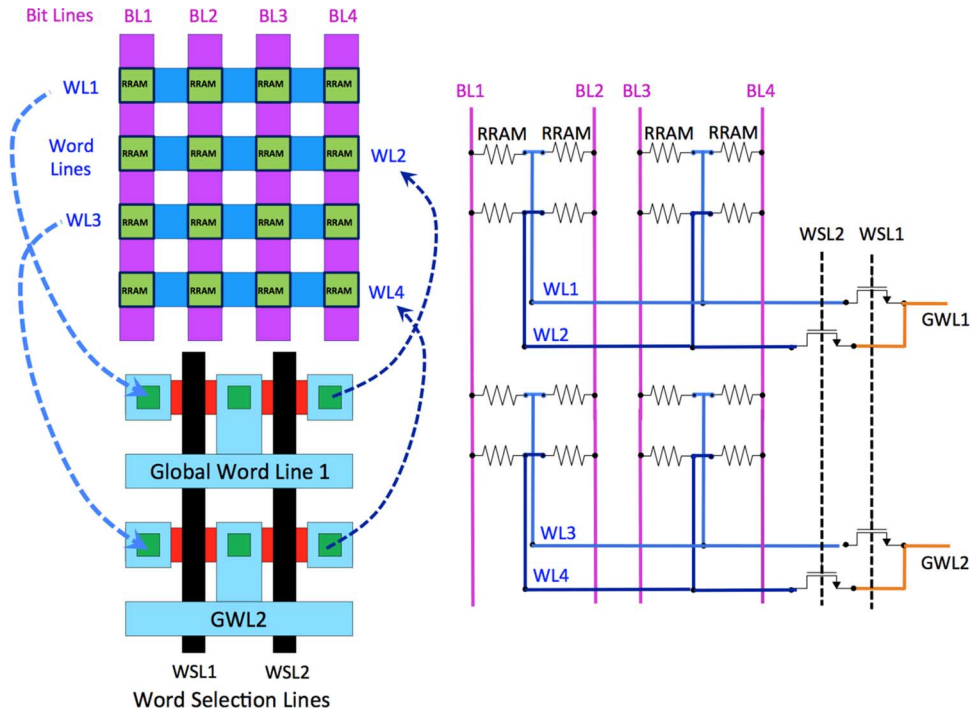


Fig. 9. 1T4R architecture layout and schematic.

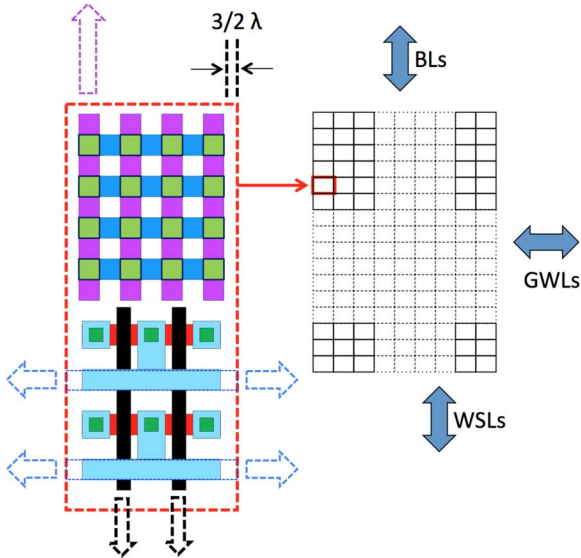


Fig. 10. Construction of a big array by repeating the 4×4 sub-arrays.

to V_g ; all bit lines and global word lines are then pre-charged to around half of the SET voltage (V_{SET}) at t_1 . Next, the selected bit line is raised to V_{SET} at t_2 . The selected global word line is connected through a compliance current source to ground (GND) at t_3 to initiate the SETting. This current source controls the SET current and the resulting low resistance (R_L). The selected RRAM cell is SET between t_3 and t_4 , at which point all bit lines and global word lines are pulled down to ground. The voltage V_{SET} is chosen to be high enough to SET the selected RRAM cell between t_3 and t_4 , while $(1/2)V_{SET}$ is low enough so as not to falsely SET other unselected cells.

Fig. 12 explains the problem of raising the selected bit line voltage to V_{SET} directly, without pre-charging it to $(1/2)V_{SET}$. As shown in this figure, in the 1T4R architecture, there are floating word lines (e.g., WL2 and WL4 in green) connecting to unselected transistors. The rise time of the voltage on these floating word lines, from ground up, depends on the resistance of the unselected RRAM cells; it could be relatively slow if all the RRAM devices connected to it are in the high resistance state. Hence, if the floating word line voltage is still low while the selected bit line voltage is rapidly pulled up to V_{SET} , the RRAM cells drawn in red would possibly be falsely SET. In contrast, with the help of pre-charging the selected bit line to only $(1/2)V_{SET}$, the unselected RRAM cells (including the ones with the floating word lines) in the array are subject to only $(1/2)V_{SET}$ or no bias, which prevents false SETting.

For the RRAM devices that require FORMing, the same procedure could be used to initially FORM the cells, except that a higher FORMing voltage is applied to the selected bit line.

Fig. 13 shows the scheme to RESET a selected RRAM cell to high resistance (R_H). The procedure is similar to SET, except that the selected bit line is grounded and the selected global word line is raised to V_{RESET} for the bipolar-switching RRAM devices. The voltage V_{RESET} is chosen to be high enough to RESET the selected RRAM cell between t_1 and t_2 , while $(1/2)V_{RESET}$ is low enough so as not to falsely RESET other unselected cells.

Fig. 14 shows the read scheme that supports the simultaneous reading of four RRAM cells connected to the selected transistor. This scheme is similar to the one described in [12]. There are four sense amplifiers (SA) at the end of the four bit lines, and each sense amplifier is a current mirror with an internal amplifier that forces the corresponding bit line voltage close to V_{SA} . As a result of also applying the same voltage V_{SA} to the un-

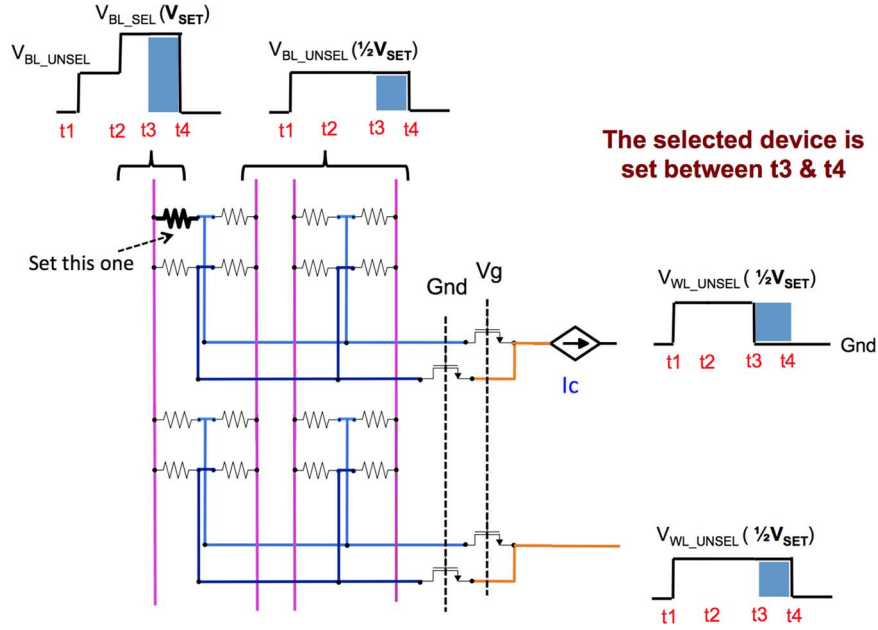


Fig. 11. SET scheme for 1T4R architecture.

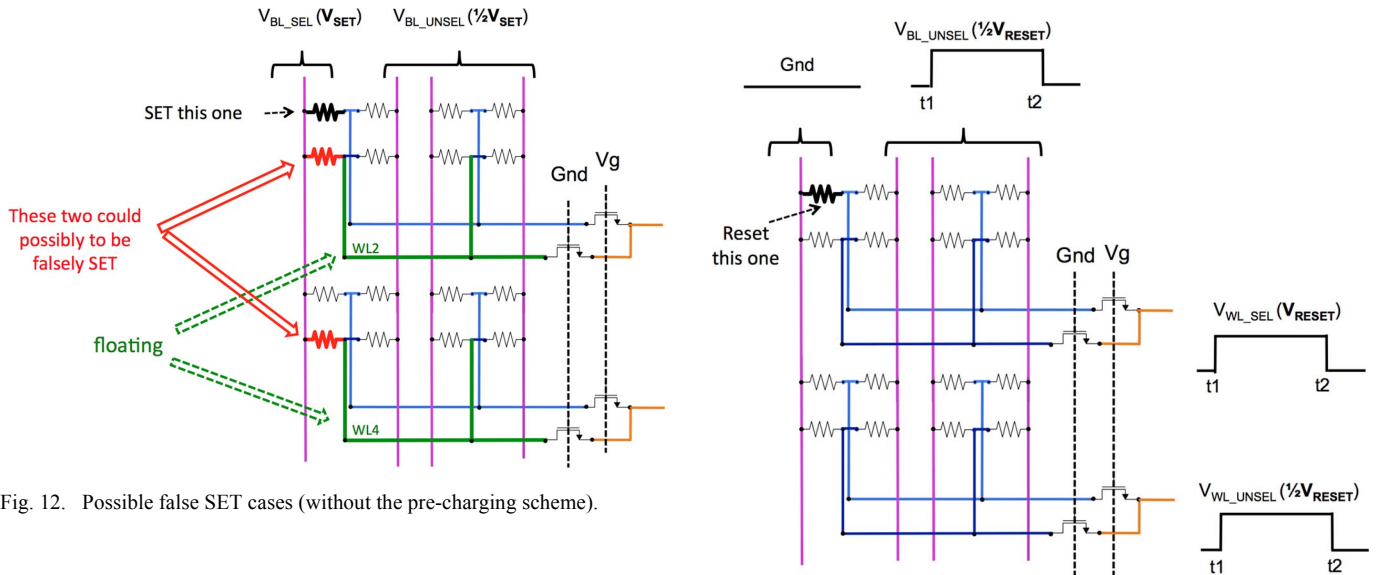


Fig. 12. Possible false SET cases (without the pre-charging scheme).

lected global word lines, the voltage across the RRAM cells connected to the unselected global word lines and thus the leakage current flowing into the sense amplifier are minimized. The selected global word line voltage V_{READ} creates a voltage difference $V_{\text{READ}} - V_{\text{SA}}$ across the selected RRAM cells, and this difference in turn generates the selected RRAM current flowing into the sense amplifier. The sense amplifier output current is monitored to determine the state of the selected RRAM cell.

C. 1TNR $4 \times N$ Sub-Array and Multi-Layer RRAM Stack

The 1T4R architecture achieves a high density RRAM cross-point array that could be partitioned into multiple sub-arrays each with 4 bit lines. According to Fig. 8, a N of 4 is adequate for RRAM device with $R_L = 100 \text{ k}\Omega$ even at $F = 8.5 \text{ nm}$ technology node.

For some designs that are based on RRAM devices with higher R_L [13], 1TNR architecture with $N > 4$ can be adopted.

Fig. 13. RESET scheme for 1T4R architecture.

1TNR cross-point sub-array looks very similar to the 1T4R sub-array; the only difference is that the former has $N \times \text{RRAM}$ devices in each row. The row number for an RRAM sub-array is still 4 so that there are total $4 \times N$ RRAM devices sandwiched between N top bit lines and four word lines.

It is possible to extend the 1TNR architecture with 3D stacking of RRAM cells. Fig. 15 shows the 3D layout of the 1T4R architecture with one and two layers of RRAM devices. For the two-layer RRAM layout, the architecture actually becomes 1T8R (one-transistor-eight-RRAM) and needs an additional 6λ to accommodate the via connecting the word line of the first layer of RRAM devices to the word line of the second layer of RRAM devices. Even though the via needs

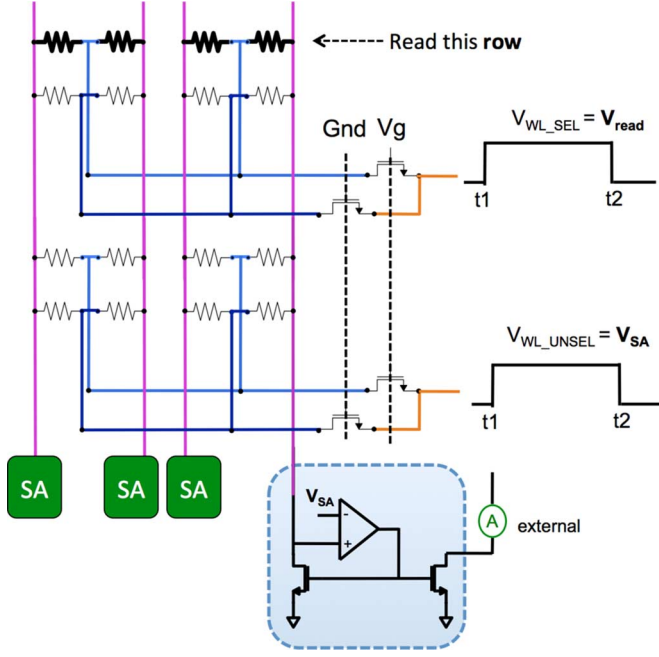


Fig. 14. Read scheme for 1T4R architecture.

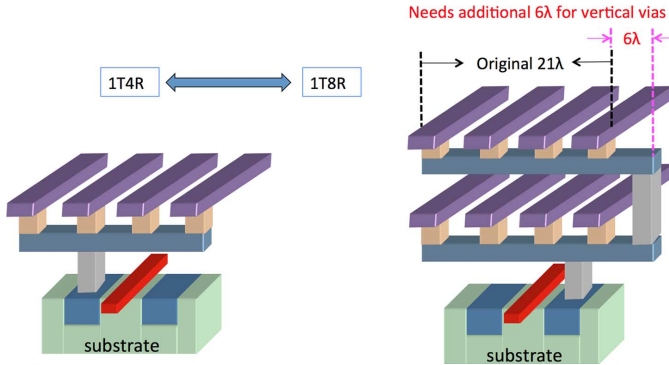


Fig. 15. Architectures with one (1T4R) and two (1T8R) layers of RRAM devices.

additional planar space, the $2.5F^2$ effective RRAM cell size of the two-layer design is smaller than that of the one-layer design.

IV. EXPERIMENTAL RESULTS

This section describes the design and fabrication of a proof-of-concept test chip with fully integrated HfO_x based RRAM devices [14]. The test chip is fabricated based on a TSMC 40 nm CMOS process and its micrograph and floor plan are shown in Fig. 16. It implements a 64×64 RRAM cross-point array and operates the schemes presented in Section III-B. It includes also the WL and BL decoders and drivers. Due to the dense dummy metal structures required to fulfill the metal density rules, no visible features can be seen. The dark rectangle region is the logo of this test chip, which is built on the top layer of metal.

A. RRAM Device Characteristics

The characteristics of the RRAM devices on the test chip are summarized as follows:

- FORMing voltage: $V_{\text{FORM}} \approx 3 \text{ V}$. $T_{\text{FORM}} \approx 1 \text{ ms}$;
- SET voltage: $V_{\text{SET}} \approx 1.5 \text{ V}$. $T_{\text{SET}} \approx 10 \mu\text{s}$;

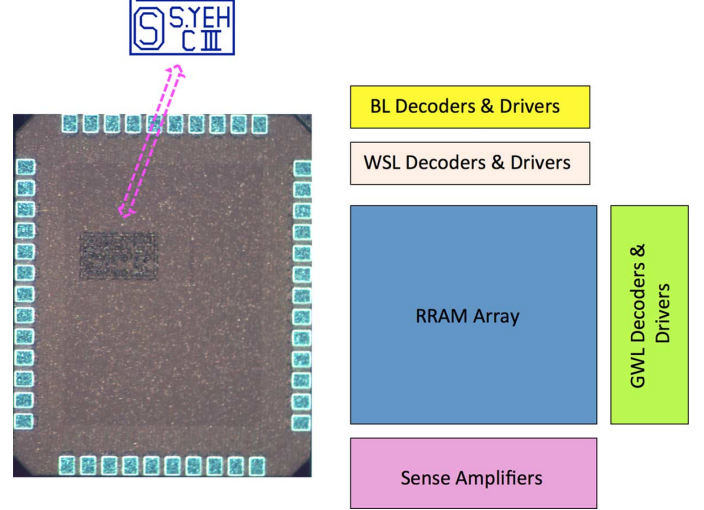


Fig. 16. 1T4R test chip micrograph and floor plan.

- RESET voltage: $V_{\text{RESET}} \approx -1.5 \text{ V}$. $T_{\text{RESET}} \approx 1 \text{ ms}$;
- high resistance state: $R_H > 350 \text{ k}\Omega$;
- low resistance state: $R_L < 40 \text{ k}\Omega$.

B. Array Programming Strategy

The cell needs significant FORMing. Therefore, tremendous care is necessary to FORM a selected RRAM device in a cross-point array without falsely SETting other unselected ones that have already been FORMed and RESET. A strategy was developed to FORM the cells one by one from top to bottom in the first column, then from left to right for each column. The key to the FORMing procedure is to RESET the selected cell to HRS after FORMing, for two reasons:

- Doing so makes the FORMing current of individual cells more accurately controlled by the compliance current source. Since all the cells that have been FORMed are RESET back to HRS, the compliance current will flow mostly to the selected cell to be FORMed.
- Not doing so will over-RESET a cell that has been FORMed, as illustrated in Fig. 17. In this figure, the RRAM cells drawn in red, blue, green, and black are FORMed and RESET, FORMed but not RESET, to be FORMed, and not yet FORMed, respectively. To FORM the selected cell the selected bit line is applied with a high voltage V_{FORM} while the unselected bit line is applied with a low voltage to prevent the false SETting of the cells drawn in red. The high V_{FORM} will create a high voltage on the floating word line 1 (WL1) that would very likely over-RESET the red cell pointed out in the figure.

Similar to the process of FORMing, but simpler, the SET/RESET flows are fairly straightforward as described in Fig. 11 and Fig. 13.

C. Sample Array Test Results

A particular attribute of the 1T4R architecture is that, by the isolation of the select transistors, a big array could be divided into a number of independent four-bit-line sub-blocks, running in parallel with each other. When programmed, one sub-block will not affect the others whose select transistors are

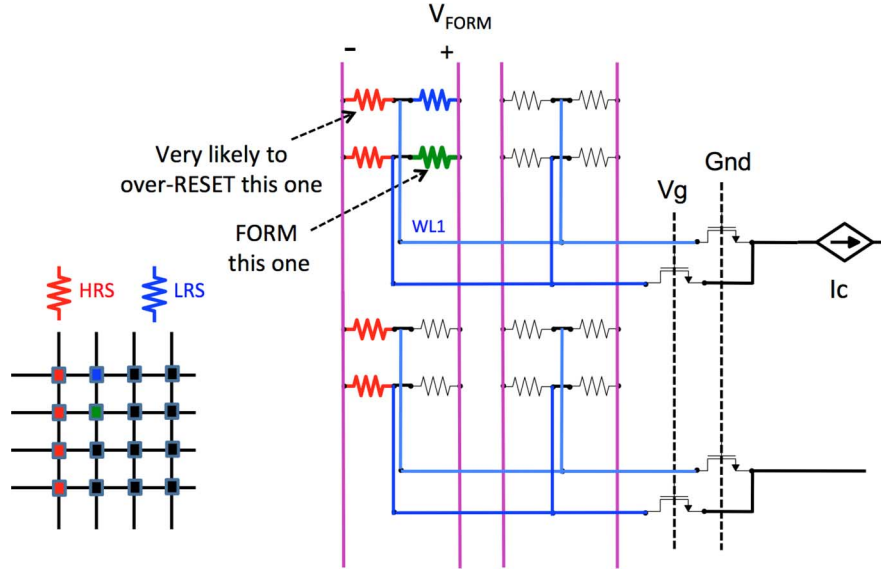


Fig. 17. Possible over-RESET case in the FORMing process.

turned off. Making use of this feature, a sample four-bit-line sub-array example is shown as a representative of the array programming result in this section. Programming of the array uses the procedures given in the previous section. The test array is first FORMed and RESET, and then programmed with specific test patterns. The array resistance is measured with a voltage of 0.5 V across the RRAM device (e.g., $V_{\text{READ}} = 1 \text{ V}$ and $V_{\text{SA}} = 0.5 \text{ V}$). Speed and power consumption are not the major concerns in the characterization of the proof-of-concept test chip; they are not the focus of the chip circuit design.

The first test is to assure that, during the SETting and RESETing of a particular cell, the other three cells connected to the same select transistors are not affected. Fig. 18 summarizes the cell resistances as the four cells in a row are SET and then RESET one at a time:

- There is no significant disturbance of neighboring cells as a cell is programmed. This demonstrates that an RRAM cell in an arbitrary location of the array can be programmed independently.
- This result also demonstrates the superiority of the 1T4R architecture, a full row of cells could be SET, one by one, with well-controlled LRS resistance.
- The one-by-one RESET result shows the device-to-device HRS resistance variation in the RESET process. Unlike the SET process's use of the compliance current to control R_L , no effective way has been found experimentally to control R_H well in the RESET process for the RRAM device in the test chip. Controlling the RESET pulse height V_{RESET} and pulse width t_{RESET} has not been effective in reducing the device-to-device R_H variation. The difficulty of controlling R_H well is possibly due to the very stochastic nature of the RRAM RESET process. In [15] the current fluctuation in the RESET process is studied. The variations of the tunneling gap distances and new oxygen vacancies generation in the ruptured conductive filament region are found to be the factors that result in the R_H variation.

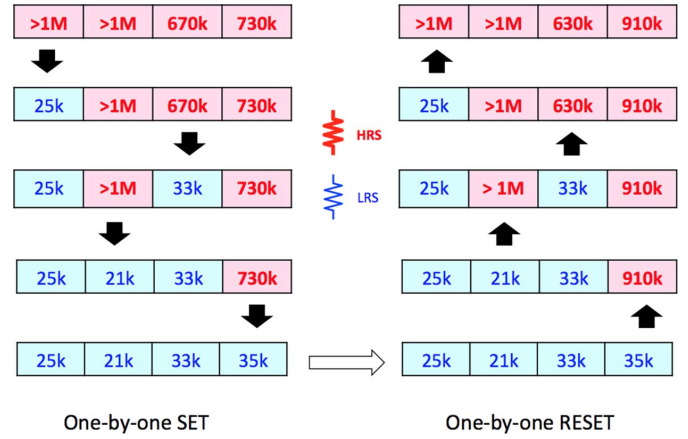


Fig. 18. SET and RESET of individual cell in a 1T4R structure.

	HRS	> 1M	21k	368k	26k
	LRS	33k	538k	18k	> 1M
		> 1M	22k	500k	31k
		18k	412k	28k	700k

$$\text{On/off ratio} = R_H/R_L > (368\text{k}/33\text{k}) = 11$$

 Fig. 19. Checkerboard patterning of 4×4 array.

Fig. 19 illustrates a 4×4 sub-array that has been programmed with a checkerboard pattern. The LRS resistance spread as shown is quite well controlled, while the HRS resistance again exhibits a noticeable variation. Even so, the lowest R_H of 368 k Ω and the highest R_L of 33 k Ω represent an on/off ratio larger than 10. The row-by-row checkerboard patterning does not induce significant disturbance to the cells in the same column. This further proves the random programmability of the 1T4R RRAM array.

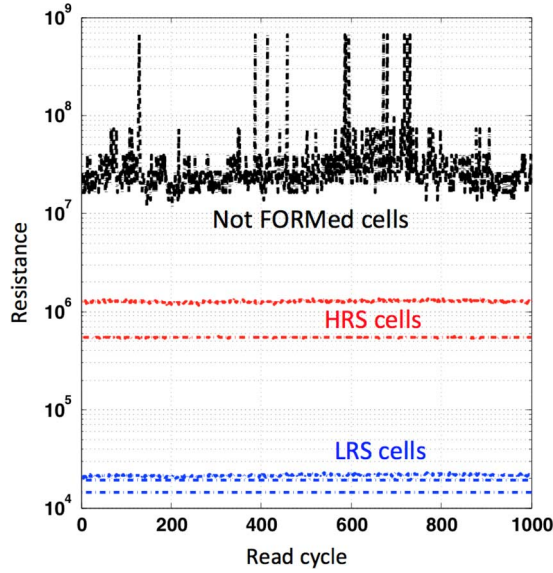


Fig. 20. Read disturbance test result.

Besides the performance of the programming schemes, another critical measure of nonvolatile memory design is the read disturbance. A read disturbance test can determine whether reading the memory cell would cause nearby cells in the same memory block to change over time.

Fig. 20 shows the resistances of a sample array of cells as they are read repeatedly for up to 1,000 cycles. These cells are in various initial states: some are fresh (not yet FORMed), some are SET into R_L , and some are RESET back to R_H . As shown in the figure, there is no measurable disturbance for cells in LRS and HRS. The variation in the resistances of fresh cells is mostly due to current resolution of the sense amplifier.

D. Endurance and Retention Test Results

This section reports on reliability of the RRAM, beginning with an endurance test. an RRAM cell was SET and RESET repeatedly for 10^5 times. Fig. 21 shows that $10\times$ on/off window was maintained.

The second test was on retention. The RRAM devices in either HRS or LRS were baked at various temperatures for 24 hours; the results are shown in Fig. 22. The X-axis shows the temperature and the Y-axis presents the RRAM resistance change ratio after baking at the specified temperature. The resistance change at 125°C and 150°C is minor. At higher temperatures, R_H increases faster than R_L , which may result in a larger on/off window.

E. Comparison With 1T1R and Cross-Point Arrays

Although we did not implement 1T1R and cross-point arrays for direct comparison with the 1TNR architecture, the following general results are expected.

- Array area: the 1TNR array has an effective cell size of $4F^2$, similar to that of cross-point array, and much smaller than that of 1T1R array with an effective cell size of about $11.7F^2$.
- Array efficiency: the 1TNR architecture requires a word select line (WSL) decoder in addition to the bit line and

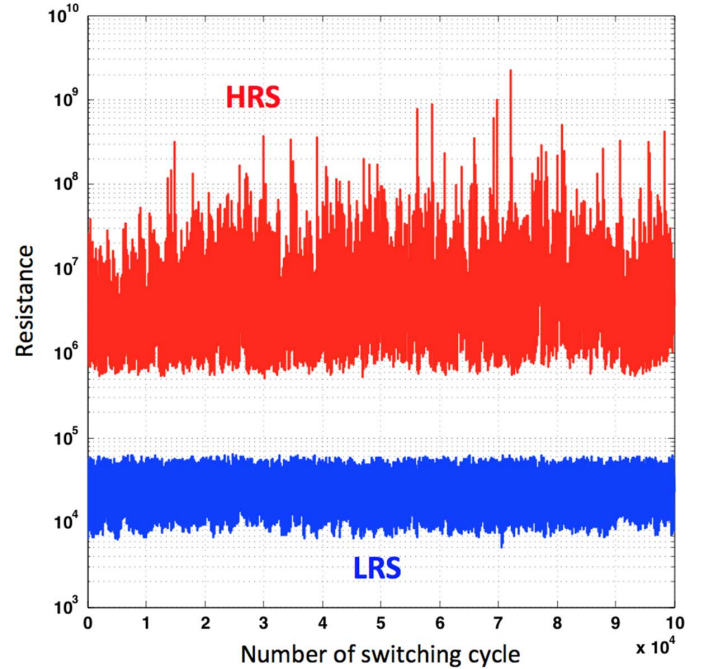


Fig. 21. Endurance test result.

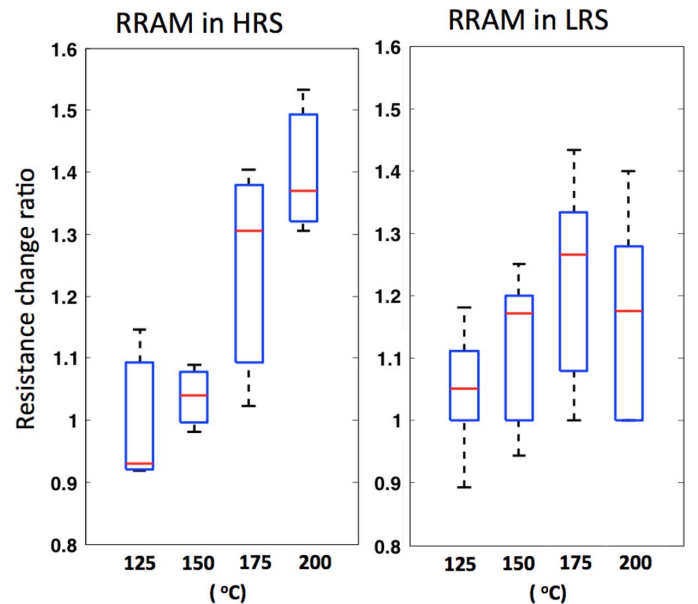


Fig. 22. Retention test result.

word line decoders in cross-point array. Furthermore, it may be possible to fold the decoders under the cross-point array. Hence, the array efficiency of 1TNR array will not be as high as that of cross-point array.

- Speed and power: the cross-point array with the highest cell density and array efficiency is expected to have the fastest read speed and lowest power. The speed and power performance of 1TNR array should approach those of cross-point array, and better than 1T1R array. The write speed will largely be determined by the programming characteristics of the RRAM cell and may not differ much for the three arrays.

- Read and write error rate: 1T1R array is expected to be the most reliable because of the isolation of individual cell during write and read. The cross-point array is expected to be the most susceptible to write and read errors due to the leakage paths. The 1TNR array offers an effective approach to control the leakage paths, write and read errors.

V. CONCLUSION

The cross-point RRAM array offers the minimum $4F^2$ RRAM cell size; however, it has the drawback of interference between RRAM cells. The work presented in this paper has sought to extend the understanding and quality of RRAM array design in two areas. The first is a thorough analysis of RRAM array performance in the presence of significant interconnect limitations. While a more advanced CMOS process would provide faster transistors and lower power consumption, it adversely worsens the interconnect properties of parasitic wire resistance and capacitance. In Section II, detailed analyses of the impact of scaled interconnect resistance on RRAM array performance are given. Compact and efficient array models are proposed to facilitate the analysis. The analysis methodology and results provide insightful references for future RRAM integration in advanced technologies. The analysis results also suggest a minimum required RRAM LRS resistance for advanced process technology. Second, a new 1TNR array architecture is proposed. It possesses the same RRAM density as the cross-point array but substantially reduces the leakage path problem. Programming and read schemes to operate the 1TNR array are presented. A 1T4R test chip is presented in Section IV. To FORM the RRAM array on the test chip, a unique FORMing strategy is presented. The measurement result validates the new 1T4R architecture.

ACKNOWLEDGMENT

The authors would like to thank the support of industrial members of the Stanford Non-volatile Memory Technology Research Initiative, and Taiwan Semiconductor Manufacturing Company.

REFERENCES

- [1] L. C. Fujino, A. Wang, and K. C. Smith, "Through the looking glass, part 2 of 2: Trend tracking for ISSCC 2013," *IEEE Solid-State Circuits Mag.*, vol. 5, pp. 71–89, 2013.
- [2] E. Ou and S. S. Wong, "Array architecture for a nonvolatile 3-dimensional cross-point resistance-change memory," *IEEE J. Solid-State Circuits*, vol. 46, no. 9, pp. 2158–2170, Sep. 2011.
- [3] S.-S. Sheu *et al.*, "A 5 ns fast write multi-level non-volatile 1 k bits RRAM memory with advance write scheme," in *Symp. VLSI Circuits Dig.*, 2009, pp. 82–83.
- [4] S.-S. Sheu *et al.*, "A 4 Mb embedded SLC resistive-RAM macro with 7.2 ns read-write random-access time and 160 ns MLC-access capability," in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, 2011, pp. 1483–1496.
- [5] A. Kawahara *et al.*, "An 8 Mb multi-layered cross-point ReRAM macro with 443 MB/s write throughput," *IEEE J. Solid-State Circuits*, vol. 48, no. 1, pp. 178–185, Jan. 2013.
- [6] *International Technology Roadmap for Semiconductors*, Semiconductor Industry Association, SEMATECH, Austin, TX, USA, 2010.

- [7] C.-W. S. Yeh, "Analysis and design of high density RRAM arrays," Ph.D. dissertation, Stanford Univ., Stanford, CA, USA, 2014.
- [8] J. Liang, S. Yeh, S. S. Wong, and H.-S. P. Wong, "Scaling challenges for the cross-point resistive memory array to Sub-10 nm node—An interconnect perspective," in *Proc. Int. Memory Workshop*, 2012, pp. 61–64.
- [9] F. T. Chen, Y.-S. Chen, T.-Y. Wu, and T.-K. Ku, "Write scheme allowing reduced LRS nonlinearity requirement in a 3D-RRAM array with selector-less 1TNR architecture," *IEEE Electron Device Lett.*, vol. 35, no. 2, pp. 223–225, 2014.
- [10] H.-Y. Chen, S. Yu, B. Gao, P. Huang, J. Kang, and H.-S. P. Wong, "HfOx based vertical resistive random access memory for cost-effective 3D cross-point architecture without cell selector," in *IEDM Tech. Dig.*, 2012, pp. 497–500.
- [11] L. Zhang, S. Cosemans, D. J. Wouters, B. Govoreanu, G. Groeseneken, and M. Jurczak, "Analysis of vertical cross-point resistive memory (VRRAM) for 3D RRAM design," in *Proc. Int. Memory Workshop*, 2013, pp. 155–158.
- [12] M. Johnson *et al.*, "512-Mb PROM with a three-dimensional array of diode/antifuse memory cells," *IEEE J. Solid-State Circuits*, vol. 38, no. 11, pp. 1920–1928, Nov. 2003.
- [13] W. Kim *et al.*, "Forming-free nitrogen-doped AlO_x RRAM with sub- μ A programming current," in *Symp. VLSI Technology Dig. Tech. Papers*, 2011, pp. 22–23.
- [14] Y. S. Chen *et al.*, "Highly scalable hafnium oxide memory with improvements of resistive distribution and read disturb immunity," in *IEDM Tech. Dig.*, 2009, pp. 105–108.
- [15] S. Yu, X. Guan, and H.-S. P. Wong, "On the switching parameter variation of metal oxide RRAM—Part II: Model corroboration and device design strategy," *IEEE Trans. Electron Devices*, vol. 59, no. 4, pp. 1183–1188, 2012.



Chih-Wei Stanley Yeh (S'10–M'12) received the B.S. (with honors) and M.S. degrees in electrical engineering from National Cheng Kung University, Tainan, Taiwan. He received the Ph.D. degree in electrical engineering from Stanford University, Stanford, CA, USA, in 2014.

From 1999 to 2001, he was with Taiwan Semiconductor Manufacturing Company (TSMC), Hsinchu, Taiwan, working as an Analog IC designer. In 2001, he joined Mediatek Inc., Hsinchu, Taiwan, where he was a technical manager. During his time with Mediatek, he participated in various projects including baseband chips, power management chips, transceiver chips, and more. He is currently with Apple Inc., Cupertino, CA, USA.

Dr. Yeh received the National Science Council (NSC) Creativity Award in 1995 and the Acer Thesis Award in 1997, for his work on Sigma-Delta ADC design.



S. Simon Wong (M'83–SM'91–F'99) received the Bachelor degrees in electrical engineering and mechanical engineering from the University of Minnesota, Minneapolis, MN, USA, and the M.S. and Ph.D. degrees in electrical engineering from the University of California, Berkeley, CA, USA.

His industrial experience includes semiconductor memory design at National Semiconductor (1978–1980) and semiconductor technology development at Hewlett Packard Labs (1980–1985). He was an Assistant Professor at Cornell University (1985–1988). Since 1988, he has been with Stanford University, Stanford, CA, USA, where he is now a Professor of electrical engineering. His current research concentrates on understanding and overcoming the factors that limit performance in devices, interconnections, on-chip components and packages. He is also on the board of Pericom Semiconductor.