Next-Generation Ultrahigh-Density 3-D Vertical Resistive Switching Memory (VRSM)—Part I: Accurate and Computationally Efficient Modeling

Shengjun Qin, Zizhen Jiang, Haitong Li, Shosuke Fujii, Dongjin Lee, S. Simon Wong, and H.-S. Philip Wong

***Abstract*— Resistive switching memory (RSM) shows potentials for high-capacity storage because of its simple cell structure, small footprint, and good scalability. This two-part article discusses how to implement ultrahigh-density (∼terabits) storage with RSM covering design considerations from device to memory array architecture. In Part I of this two-part article, an accurate and computationally efficient model is developed to study 3-D vertical RSM (VRSM). In this article, we use 3-D VRSM with a hexagon-patterned pillar layout (3-D hexagon VRSM) as an example to elaborate our approach. Using the parasitic resistance extracted from physics-based 2-D field solver as the reference, we develop a lumped resistor network for SPICE simulation to accurately capture all leakage currents of the array. This full resistor network is further simplified to a reduced network to achieve high computational efficiency and maintain the full network accuracy (with a relative error <2%). Without this simplification, practical memory array sizes cannot be simulated efficiently. The interplay between the memory cell resistance values and selector nonlinearity (NL) on the maximum array size and the write/read margins under the worst case scenarios is discussed. Large memory cell resistance (low-resistance states ≥100 k) with enough NL (∼10^3) from the selector is recommended for successful write/read operations in a single 3-D hexagonal VRSM array of megabit scale.**

***Index Terms*— 3-D, array, conductive-bridging random access memory (CBRAM), nonlinearity (NL), one selector one resistor (1S1R), phase-change memory (PCM), resistive random access memory (RRAM), resistive switching memory (RSM), selector, SPICE.**

I. Introduction

RESISTIVE switching memories (RSMs), including resistive random access memory (RRAM) [1], conductive-bridging random access memory (CBRAM) [2], and phase-change memory (PCM) [3], are nonvolatile memories that rely on different resistance states to denote data stored in each cell. RSMs remain strong candidates for high-density, low-cost, and low-power nonvolatile data storage in the sub-20-nm technology regime [1]–[4] due to multiple advantages of fast programming speed (∼ns) [5], small cell size ($4F^2$) and good scalability [6], and fabrication simplicity and CMOS compatibility [7], [8]. However, common drawbacks of passive RSM array circuits are the nontrivial leakage currents from half- and unselected cells and interconnect wires, which significantly degrades the write/read margins and the achievable maximum array size. Therefore, two-terminal selecting devices (selectors) with good nonlinearity (NL) integrated with the RSM cells are essential in RSM arrays to minimize leakage currents. Substantial studies have been done at the device level, and various types of selectors have been reported [9]–[13]. Nevertheless, device-level characterizations of one resistor (1R) and one selector one resistor (1S1R) are not adequate for accurate estimation of leakage currents in large arrays. Full array analysis with all parasitic resistances captured is necessary for the exploration of selector requirements that attains enough write/read margin. We note that the selector in series with the memory cell is an integrated device. As such, the NL of the selectors must be defined based on explicit definition of low-resistance state (LRS) and unselected resistance state (URS) of 1S1R cells in the RSM array analysis. Defining NL for a selector alone without identifying the LRS and URS of the 1S1R cells leads to ambiguities in the discussions of selector requirements for RSM arrays.

The 3-D NAND Flash is the current mainstream nonvolatile memory technology that achieves terabit capacity [14]–[17].

Manuscript received August 14, 2019; revised October 3, 2019; accepted October 19, 2019. Date of publication November 18, 2019; date of current version November 27, 2019. This work was supported in part by the member companies of the Stanford Non-Volatile Memory Technology Research Initiative (NMTRI) and in part by the Samsung Electronics' University R&D Program. The review of this article was arranged by Editor J. Kang. *(Corresponding author: Zizhen Jiang.)*

S. Qin, Z. Jiang, H. Li, S. S. Wong, and H.-S. P. Wong are with the Department of Electrical Engineering and Stanford SystemX Alliance, Stanford University, Stanford, CA 94305 USA (e-mail: sjqin@stanford.edu; jiangzz@stanford.edu; hspwong@stanford.edu).

S. Fujii is with the Department of Electrical Engineering and Stanford SystemX Alliance, Stanford University, Stanford, CA 94305 USA, and also with Kioxia Corporation, Yokkaichi 512-8550, Japan.

D. Lee is with the Department of Electrical Engineering and Stanford SystemX Alliance, Stanford University, Stanford, CA 94305 USA, and also with the Samsung Semiconductor R&D Center, Hwaseong 18848, South Korea.

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

Digital Object Identifier 10.1109/TED.2019.2950606

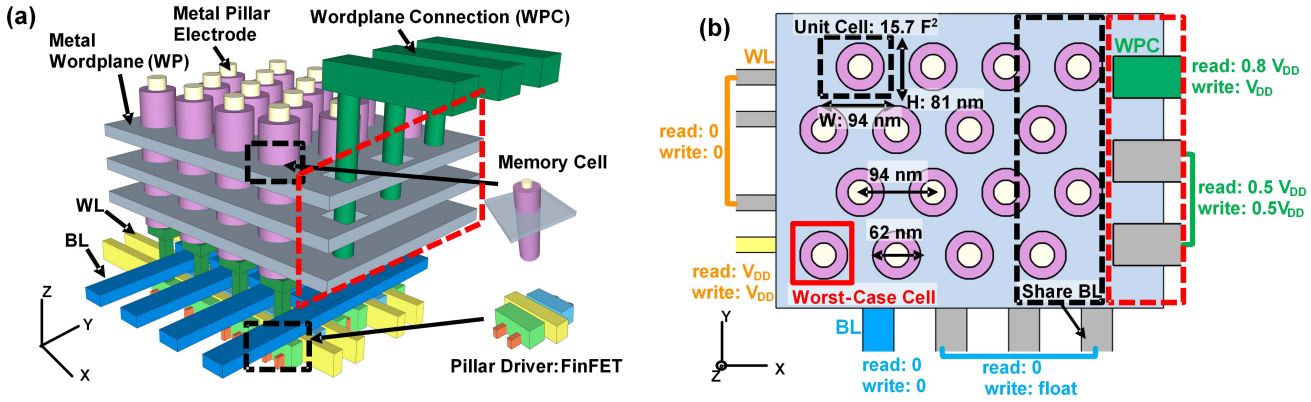

Fig. 1. Schematics of (a) side view and (b) top view of 3-D VRSM with compact staircase connections.

Similarly, 3-D RSM can also take advantage of the third dimension by stacking multiple layers of memory cells [18]–[36]. There are two broad classes of 3-D RSM. The first option is a 3-D horizontal RSM (HRSM) by stacking up multiple 2-D cross-point arrays [18]–[20]. The fabrication cost of 3-D HRSM increases with the number of layers in the third dimension, as lithography steps and masks are required by each layer of the 2-D cross-point array. The other option is 3-D vertical RSM (VRSM) that uses far less lithography steps in fabrication and thus has a lower cost per bit compared to 3-D HRSM [21]. One typical 3-D VRSM structure studied in [22]–[33] only requires one lithography mask to pattern metal plane layers in the third dimension. In [29], we discussed this type of 3-D VRSM using patterned pillars with a hexagonal layout, resembling the pillar layout in today's 3-D NAND Flash [37]–[39]. This hexagonal layout achieves higher pillar density with the same pillar diameter and pillar-to-pillar pitch. Another type of 3-D VRSM is 3-D comb VRSM with double cell density per pillar, which has been studied in [30] and [31]. In order to determine which type of 3-D VRSM is the best candidate for ultrahigh-density storage, accurate modeling for both hexagon and comb 3-D VRSM structures is required.

This two-part article focuses on exploring 3-D VRSM as candidates for ultrahigh-density data storage: In Part I of this article, we first investigate how to accurately model the 3-D hexagon VRSM with explicit and proper definitions of LRS, URS, and NL. NL requirements and the impact of memory cell resistance values on achievable array sizes are discussed for the 3-D hexagon VRSM at the single array level. Our modeling approach can also be utilized for other types of 3-D VRSM arrays. In Part II of this article [42], we further provide design guidelines from device to architecture levels to store at least 1 Tb/chip, with an even higher bit density (for single bit per cell) compared with the most advanced 3-D NAND Flash (with 3 bits/cell). The 3-D hexagon VRSM and 3-D comb VRSM are also compared in Part II.

II. 3-D VRSM Structure

Schematics of our 3-D VRSM structure are shown in Fig. 1. In order to achieve the maximum areal density in a single array, we adopt a hexagonal layout for the pillar array to realize the highest density for the same pillar diameters and distance between the pillars. The 1S1R (or 1R) cells are sandwiched between metal core pillars and metal wordplanes (WPs). The inner metal pillar at the center serves as one electrode of 1S1R (or 1R) cells connecting with the drain of a pillar driver (or access transistor) at the bottom. The outer diameter of the pillar is limited by lithography and the choice of pillar drivers (discussed in Section II-A of Part II [42]). For our case of using 7-nm FinFET (with double-fin) as pillar drivers, the dimension along the y-axis is limited by the contacted gate pitch (54 nm) [40]. In this article, we use the regular hexagonal pillar layout for simplicity, and thus, the dimension along the x-axis with the FinFET isolation is larger than the minimum value. With all these considerations, we choose a practical size for the pillars with a diameter of 62 nm and have the pillar-to-pillar pitch of 94 nm (a unit cell of $15.7F^2$), which leaves enough room to allow further optimization of the areal density to approach the minimum FinFET isolation dimensions in the x-direction. WPs serve as the other electrode of 1S1R (or 1R) cells. We adopt a compact staircase (marked in the red dashed box in Fig. 1) to map the WP connections (WPCs) for accessing each metal WP in this array structure, which reduces the area of staircase compared to traditional staircases. (More details of the array dimensions and layout are covered in Part II [42].) As the voltage bias is applied at the top-right corner on the top WP in Fig. 1(b), the diagonal cell on the same WP is the worst case cell (marked with the red solid box), which has the least voltage drops across the memory cell. Our analyses are conducted for this worst case scenario.

The access transistors underneath pillars (pillar drivers) need to provide sufficient drive current for programming operations of 1S1R (or 1R) cells. The gates of transistors in the same row along the x-axis are chained together as the wordline (WL) and the sources of transistors in the same column along the y-axis are connected as the bitline (BL), while the drains of the transistors are affixed to each core metal pillar. This structure has advantages, such as high pillar areal density and low fabrication cost. The detailed discussions of pillar driver dimensions and transistor candidates for achieving enough saturation current and good areal density are also included in Part II [42]. In this article, we focus on the methodology to

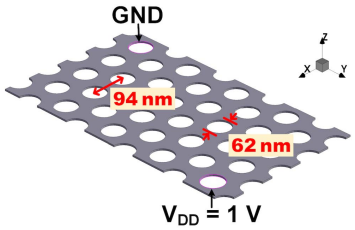

Fig. 2. Illustration of a single metal plane structure with 32 MHs simulated in the 2-D field solver (corresponding to a 4 × 8 single-layer array). Two contacts are defined at two diagonal corners. Metal plane material is copper.

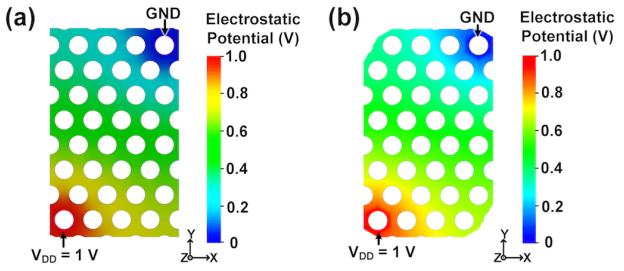

Fig. 3. Electrostatic potentials extracted from (a) 2-D field solver and (b) resistor network simulation in SPICE. Voltage biases (=1 V) at the bottom-left corner.

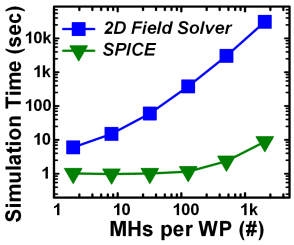

Fig. 4. Comparison of simulation time for a single metal plane with 2-D field solver and resistor network in SPICE. The simulation time with 2-D field solver reaches up to about 9 h for 2k MHs.

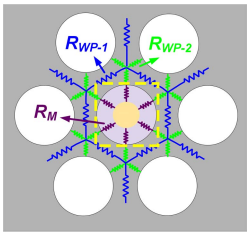

Fig. 5. Top view of a unit in the constructed resistor network. $R_{WP\text{-}1}$ and $R_{WP\text{-}2}$ are parasitic resistances on WPs. Six equivalent R_M represents one 1S1R cell surrounded by $R_{WP\text{-}1}$ and $R_{WP\text{-}2}$.

model this resistor network in an accurate and computationally efficient manner.

III. Ground Truth—2-D Field Solver

To understand the impact of parasitic leakage current on the write/read margin of the 3-D VRSM, it is necessary to accurately capture leakage current paths, especially for modeling the distributed parasitic resistance of metal planes with the hexagonal pillar layout. We first study these metal planes with physics-based 2-D field solver (2-D simulation in Sentaurus [43]). The structure that we used to simulate a single metal plane in the 2-D field solver is shown with dimensions of pillar diameter and pillar pitch in Fig. 2. In order to obtain current paths throughout this entire metal plane, we define interfaces between metal planes and pillars as contacts with equipotential around the pillars. Bias voltages are applied at the two corner contacts along the diagonal of the metal plane, corresponding to the worst case scenario. Here, in Fig. 2, memory holes (MHs) are shown as vacuum with infinite resistivity. Simulation reported here are for MHs filled with different resistivities to cover the resistance range of both LRS and URS from lower limit of LRS value (∼1 kΩ) to upper limit of URS (∼infinite).

The electrostatic potential distribution obtained from our simulation with 2-D field solver in Fig. 3(a) (an array size of 4 × 8 with 32 MHs) reveals current paths along the midpoints between two adjacent pillars when resistivities of pillar filling materials are at least 10× higher than the resistivity of the metal plane. In Fig. 3(b), we show matching electrostatic potential distribution of the resistor network from SPICE simulation, which will be discussed in Section IV-A. The resistance of the entire metal plane, which is obtained by extracting resistance between the two contacts, increases with the size of this metal plane (defined by the number of patterned MHs). However, simulation time for a single-layer metal plane with the 2-D field solver also goes up with the size of this metal plane, as shown in Fig. 4. For a single layer of 2k MHs, the simulation time is already ∼9 h with our computational platform (CPU: Intel Xeon E5-2430v2 2.5 GHz and RAM: 96 GB). Further increasing the pillar array size or stacking up more layers will consume even more simulation time and computation resources. Simulation with a 2-D field solver is impractical for large arrays with multiple layers. However, these physics-based results (metal plane resistance, metal plane electrostatic potential, and so on) can be used as references for other simulation methods.

IV. Resistor Network

With simulated resistance values from 2-D field solver, we build a lumped SPICE-based (we use HSPICE Version H-2013.03-SP2 in this article) resistor network that includes 1S1R resistance, key parasitic resistances (R_{WP-1}, R_{WP-2}, R_P, R_{BL}, and R_{WL}), and pillar drivers in the 3-D VRSM. This resistor network captures leakage path currents with high accuracy and substantially reduces simulation time and memory usage compared to the 2-D field solver.

A. Metal WPs

Current steers around pillars on the metal WPs as indicated by simulation results from the 2-D field solver [see Fig. 3(a)], especially when the resistivity of the pillar filling material is much higher than that of metal planes. Thus, we model parasitic resistances of WP in such way that the resistors R_{WP-1} (blue resistors) connect with each other around pillars

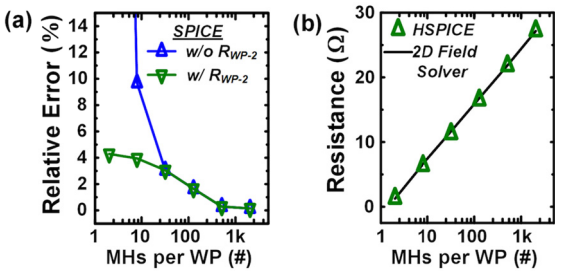

Fig. 6. (a) Importance of adding R_{WP-2} to mitigate relative error of metal plane resistance ($R_{SPICE}-R_{2DFieldSolver}$)/$R_{2DFieldSolver}$ for small arrays. With R_{WP-2}, relative error drops to <5%. (b) Metal plane resistance is accurately modeled by both R_{WP-1} and R_{WP-2}. An agreement of extracted metal plane resistance has been reached between SPICE and 2-D field solver simulation. Note that resistance values in (b) are simulated with tungsten sheet resistivity (2.75 Ω/sq) as an example.

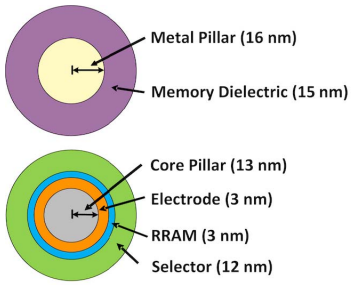

Fig. 7. Details of the layer composition of pillars. Top: memory cell only. Bottom: memory cell with selector. The thicknesses of RRAM and selector dielectrics are typical minimum practical values.

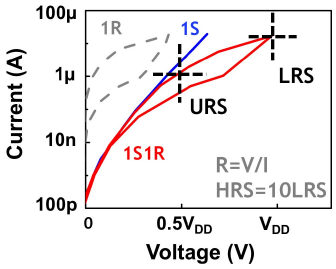

Fig. 8. Definition of LRS and URS. The gray curve is a typical RRAM dc characteristic in the SET operation. With a selector in series (blue curve), 1S1R characteristic (red curve) shows that 1S1R cell is not accessible until voltage biases above certain threshold value. For unselected cells, voltage drops on these cells are smaller than 0.5 V_{DD}, and thus, URS is defined as the equivalent resistance at 0.5 V_{DD}.

in a hexagonal pattern shown in Fig. 5. The value of R_{WP-1} and R_{WP-2} is determined by matching the results of resistor network in SPICE and the results from 2-D field solver, both simulating the resistance of an entire metal WP. A check has been done to verify that R_{WP-1} value falls in the range of hand calculated resistance with similar dimensions and the same resistivity. R_{WP-2} is necessary to be included here to correctly capture the currents between the adjacent cells [see Fig. 6(a)]. The parasitic resistance on metal WPs is accurately modeled with RWP-1 and RWP-2 for different array sizes (defined by the number of MHs per WP), shown as the green curve in Fig. 6(a). Metal plane resistances that are extracted from matching resistor network simulation in SPICE are consistent with the results from the 2-D field solver [see Fig. 6(b)]. The electrostatic potential distribution of the metal WP (with 8×8 MHs) extracted from our resistor network is shown in Fig. 3(b). The effectiveness and accuracy of modeling metal WPs with the resistor network are further validated by the comparison to the electrostatic potential distribution from the 2-D field solver in Fig. 3(a).

B. 1S1R Cells

In our simulation, we bundle a 1S1R cell together and simplify it as a voltage-dependent resistor, as shown in Fig. 7, with a detailed breakdown of 1S1R layers with specific dimensions. Layer dimensions of memory and selector cells are determined by the minimum thicknesses that we can have for each functional layer [18], [44]–[46]. Here, we use these minimum thicknesses from the reported experiments to demonstrate the feasibility to realize the small pillar dimension and pillar pitch for 3-D VRSM. For 3-D VRSM, self-rectifying RSM cells are recommended to reduce the thickness of 1S1R [18], [47]–[49].

A typical LRS value used in this article is 1 MΩ [47]–[49] and guarantees that resistance values of 1S1R cells satisfy the condition that 1S1R cell resistance is much larger than that of the metal WP. As shown in Fig. 3, the difference in electrostatic potential along the pillar outer edges is negligible in this case (the most significant difference is less than 0.1 V). Thus, 1S1R resistance can be divided into six components of equal resistance with one-sixth of 1S1R resistance value (shown as purple resistors R_M in Fig. 5). The characteristics of 1S1R are shown as the red curve in Fig. 8. Here, LRS is defined as the equivalent resistance of 1S1R at V_{DD}. With the definition of URS as the equivalent resistance of 1S1R at 0.5 V_{DD}, NL is defined as the ratio of URS and LRS. NL is independent of supply voltage when defined in this manner, and the requirement for NL holds for any supply voltage (V_{DD}). Note that NL must be defined for the combined 1S1R. NL for the selector alone (as it is sometimes referred to in the literature [10], [13]) does not adequately capture the essence of 1S1R designs. Also, note that the simplified 1S1R voltage-dependent model used in this resistor network can be replaced with a Verilog-A model with full 1S1R characteristics extracted from the experimental data.

C. Metal Core Pillars, R_P

Each inner metal core pillar connects all 1S1R cells stacking vertically along it. Current flows from each layer passing through this core pillar and eventually flows into the drain of the pillar driver. R_P is introduced to capture the core pillar resistance between the adjacent vertical layers, shown in yellow in Fig. 9. Each layer has a corresponding R_P to connect it with the layer above and below it.

D. Pillar Drivers, R_{BL}, and R_{WL}

We assume that pillar drivers in this simulation can provide enough drive current for all 1S1R cells on the same pillar and this can be achieved by a good selection among transistor candidates (discussion in Part II). R_{BL} captures wire resistance

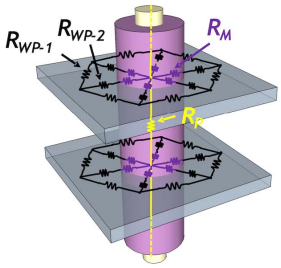

Fig. 9. Unit of full resistor network. Horizontal resistor network on metal planes are connected by pillar resistors R_P. Full resistor network for large-scale 3-D VRSM can be achieved by simply repeating this unit.

TABLE I
RESISTANCE AND RESISTIVITY USED IN THIS SIMULATION

Symbol	Description	Resistance (Ω)	Resistivity (Ω·μm)
R_{WP-1}	Parasitic resistance on WP	54.4	0.2
R_{WP-2}	Parasitic resistance on WP	27.2	0.2
R_P	Pillar resistance between two WPs	0.57	0.042
R_{BL}	Parasitic resistance on BL connecting sources of pillar drivers	1.3	0.035
R_{WL}	Parasitic resistance on WL connecting gates of pillar drivers	7.5	0.064

connecting the drains of two adjacent transistors in the same column (y-axis), whereas R_{WL} represents the wire resistance connecting two adjacent gates in the same row (x-axis) (axes refer to Fig. 1).

E. Full Resistor Network

A unit of full resistor network is established by capturing parasitic resistances discussed earlier, as shown in Fig. 9. The 3-D VRSM array structure is thus constructed by simply repeating this unit. Analyses for this 3-D VRSM are conducted under the worst case scenario with all leakage current captured. Resistances and resistivities used in this simulation are listed in Table I. Simulation results will be discussed in Section VI.

V. Reduced Resistor Network

The full resistor network mentioned earlier accurately captures the parasitic resistances in the 3-D VRSM array and is much more computationally efficient compared with the 2-D field solver. Simulation time for 2k MHs reduces to less than 10 s with the full resistor network in SPICE (green curve in Fig. 4), compared to 9 h with the 2-D field solver. However, it is still relatively time-consuming and computation-heavy for megabit scale arrays. For example, simulating an array size of 1 Mb takes up to 6 h and already reaches the upper limit of our computational capacity. Therefore, a reduced resistor network is developed to further improve simulation efficiency. For analyses targeted at the worst case scenario, the top WP where the worst case cell is located is the critical layer. All the unselected WPs are biased at only half V_{DD}, whereas the selected WP is biased at V_{DD}. Voltage drops on the unselected WPs are relatively small compared to the selected WP, which leads to the situation that the voltage distributions on these

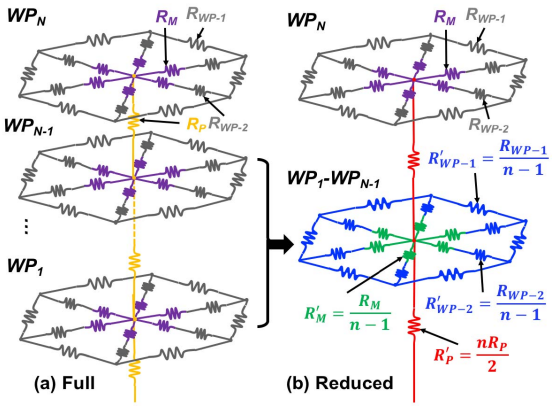

Fig. 10. (a) Full resistor network can be reduced to (b) two-layer network by collapsing noncritical $n-1$ unselected layers into one, in which resistances on metal planes (R'_{WP-1}, R'_{WP-2}, and R'_M) are $1/(n-1)$ times of original values and pillar resistance R'_P is $n/2$ of original R_P.

unselected layers are relatively uniform. Therefore, we can collapse all the noncritical, unselected $n-1$ layers (n is # of WPs and the nth layer is into one equivalent unselected WP with parasitic resistances on the metal plane and unselected 1S1R resistance being $1/(n-1)$ of the original values). R_P along the core pillar is reduced to $n/2$ times of the original R_P to maintain the same total R_P resistance seen by the circuit. The full resistor network is thus further simplified to a reduced circuit model in this manner, as shown in Fig. 10. This two-layer reduced resistor network is applicable to both write and read operations of 3-D VRSM with any WP layer size and enables computationally efficient simulation of the array size of ~Mb. The simulation time for a 1-Mb array reduces to only 2 min, and the relative errors of this reduced resistor network are verified to be <2% for submegabit arrays in Fig. 11. The maximum array size of reduced resistor network simulated is limited by the capacity of RAM (for example, the maximum array size simulated by CPU with 96 GB RAM is about 16 Mb). As the reduced network accurately reflects the full network, all the results discussed in Section VI are simulated with the reduced network.

VI. Results

We adopt the 1/2 V_{DD} biasing scheme in this simulation in which unselected WPs are biased at 1/2 V_{DD}. The selected WP is biased at V_{DD} in write operations to ensure enough voltage drop across the selected cell, whereas for read operations, the selected WP is biased at 0.8 V_{DD} to turn on the selector and at the same time avoid disturbing the original stored data. Biasing conditions for WLs and BLs are the same for both write and read operations to turn on the corresponding pillar drivers. Note that read operation collects current from all BLs simultaneously to improve read throughput (multibit read). A summary of biasing schemes and criteria for write/read operations and related data patterns are listed in Table II. Please also note that we use the same write/read criteria for both single array simulation discussed in the following and chip architecture simulation in Part II [42]. In this simulation, we set the write criteria as the voltage drop across the worst

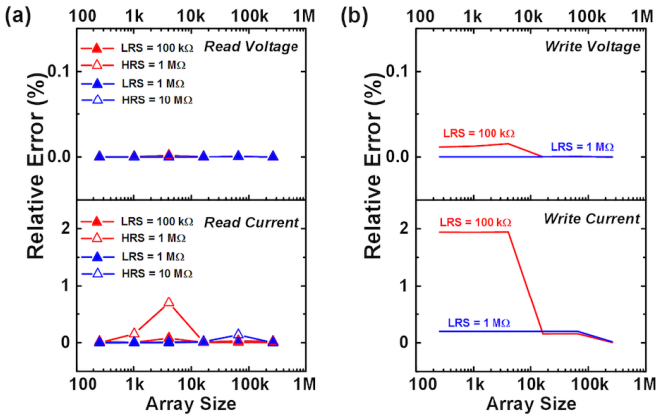

Fig. 11. Relative errors of reduced resistor network compared with full network: $(X_{\text{Full}}-X_{\text{Reduced}})/X_{\text{Full}}$, X is (a) read voltage/current and (b) write voltage/current on worst case cell. For LRS = 100 kΩ and 1 MΩ, relative errors of reduced network remain at a low level of less than 2% for an array size up to 64 × 128 × 128.

TABLE II
BIAS CONDITIONS, 1S1R RESISTANCES, AND WRITE/READ CRITERIA

Parameter	Description	Unit	Write	Read
V_{WPs}	Applied voltage on selected WP	V	V_{DD}	$0.8V_{DD}$
V_{WPu}	Applied voltage on unselected WP	V	$0.5V_{DD}$	$0.5V_{DD}$
V_{WLs}	Applied voltage on selected WL	V	V_{DD}	V_{DD}
V_{WLu}	Applied voltage on unselected WL	V	0	0
V_{BLs}	Applied voltage on selected BL	V	0	0
V_{BLu}	Applied voltage on unselected BL	V	Float	Float
R_{Ms}	equivalent resistance of selected 1S1R	Ω	LRS	Read LRS: LRS Read HRS: HRS
R_{Mu}	equivalent resistance of unselected 1S1R	Ω	URS	URS

Write criteria: 1) $V_{WC} > 0.9V_{DD}$; 2) Pillar drivers provide enough current. Read criteria: 1) $V_{WC} > 0.65V_{DD}$; 2) $I_{LRS} > 100$ nA; 3) $I_{LRS} > 5I_{HRS}$. Applied voltages for the worst-case cell analysis is shown in Fig. 1b. V_{WC} is the voltage drop on the worst-case cell and I_{LRS} and I_{HRS} are the currents through the pillar driver when reading the worst-case cell at LRS and HRS respectively.

case cell to be larger than 90% of V_{DD} (write margin >90%) when writing LRS into the worst case cell. In the read operation, we extract the voltage on the worst case cell at LRS and HRS, respectively, and calculate the read margin as the voltage on the worst case cell over the supply read voltage (0.8 V_{DD}). To obtain data for read currents, we extract I_{LRS} and I_{HRS} at $V_{WPu} = 0$ V to eliminate the current contributions from unselected WP when $V_{WPu} = 0.5$ V, and we keep the other bias conditions as the same in Table II. We require a 65% V_{DD} voltage drop on the worst case cell, $I_{LRS} > 100$ nA and $I_{LRS} > 5I_{HRS}$, for a successful read operation. All the results discussed next are simulated for single 3-D VRSM arrays with 64 vertical layers of metal WPs and memory cells with HRS/LRS = 10.

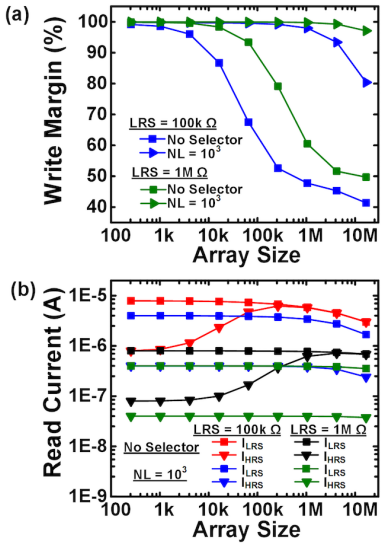

Fig. 12. Comparison between 1R array with no selector and 1S1R array with NL = 10^3, for LRS = 100 kΩ and 1 MΩ. (a) Write margin. (b) Read margin. NL is required for large arrays in both write/read operations.

For 3-D VRSM arrays with LRS = 1 MΩ, write and read margins of 1R array with no selectors drop below our criteria when the array size is above 64 kb in Fig. 12. In comparison, NL (use NL = 10^3 as an example) from selectors obviously improves the achievable array size by maintaining both write and read margins for larger array sizes up to 16 Mb. Thus, NL is clearly a requirement for large-scale arrays. As both write and read margins will benefit from the selector NL value of 10^3, however, limited benefits will be obtained from further increasing NL beyond 10^3 for the case of LRS = 1 MΩ. We also compare NL versus LRS requirements in Fig. 12. For arrays with lower LRS values, NL requirement increases accordingly. The results in Fig. 12 show that arrays with large LRS = 1 MΩ can adopt selectors with NL = 10^3 to maintain adequate write/read margin, whereas for arrays with LRS = 100 kΩ, NL = 10^3 is not enough for array sizes beyond 4 Mb. Both I_{LRS} and I_{HRS} in the case of LRS = 1 MΩ are roughly one order of magnitude lower, which results in lower energy consumption with a tradeoff of readout time. This is another benefit that we can obtain with large LRS as long as it still meets $I_{LRS} > 100$ nA and read latency requirements.

VII. CONCLUSION

We design a novel 3-D VRSM array with hexagonal pillar array layout and present a methodology to achieve an accurate and computationally efficient model for this array in this Part I of a two-part article. We build a full resistor network for the distributed resistances of the array to capture leakage currents. The resistor network has less than 5% relative errors of metal plane resistance compared to physics-based 2-D field solver simulation results. We further simplify this resistor network to a reduced network, which substantially improves computational efficiency. This reduced resistor network also retains high accuracy with only less than 2% of relative errors compared to the full resistor network. The interplay between the

selector NL and LRS, on the one hand, and the maximum array size, on the other hand, is studied with our reduced resistor network. Large LRS ($\geq$100 kΩ) with a moderate NL ($= 10^3$ as an example here) is recommended for single arrays targeting at megabit scale. This modeling approach can also be applied to other types of 3-D VRSM. To achieve high bit density for 3-D VRSM, innovations in chip architecture designs (discussed in Part II), WPs with low resistivity, and large LRS are needed.

Acknowledgment

The authors would like to thank T. Maeda from Kioxia Corporation, A. V.-Y. Thean from the National University of Singapore, H.-Y. Chen from GigaDevice Semiconductor, L. Liyanage from Western Digital, G. Hills from the Massachusetts Institute of Technology, C.-S. Lee from Apple Inc., and K.-M. Persson from Lund University and C. M. Neumann from Stanford University for the discussions.

References

[1] H. S. P. Wong *et al.*, "Metal oxide RRAM," *Proc. IEEE*, vol. 100, no. 6, pp. 1951–1970, Jun. 2012.

[2] M. Kund *et al.*, "Conductive bridging RAM (CBRAM): An emerging non-volatile memory technology scalable to sub 20 nm," in *IEDM Tech. Dig.*, Dec. 2005, pp. 754–757.

[3] H.-S. P. Wong *et al.*, "Phase change memory," *Proc. IEEE*, vol. 98, no. 12, pp. 2201–2227, Dec. 2010.

[4] R. Waser, R. Dittmann, G. Staikov, and K. Szot, "Redox-based resistive switching memories–nanoionic mechanisms, prospects, and challenges," *Adv. Mater.*, vol. 21, pp. 2632–2663, Jul. 2009.

[5] H. Y. Lee *et al.*, "Evidence and solution of over-RESET problem for HfO_X based resistive memory with sub-ns switching speed and high endurance," in *IEDM Tech. Dig.*, Dec. 2010, pp. 460–463, doi: 10.1109/IEDM.2010.5703395.

[6] B. Govoreanu *et al.*, "$10\times10nm^2$ Hf/HfO_X crossbar resistive RAM with excellent performance, reliability and low-energy operation," in *IEDM Tech. Dig.*, Dec. 2011, pp. 729–732, doi: 10.1109/IEDM.2011.6131652.

[7] M.-J. Lee *et al.*, "A fast, high-endurance and scalable non-volatile memory device made from asymmetric Ta_2O_{5-X}/TaO_{2-X} bilayer structures," *Nat. Mater.*, vol. 10, pp. 625–630, Jul. 2011, doi: 10.1038/nmat3070.

[8] W. Kim *et al.*, "Forming-free nitrogen-doped AlO_XRRAM with sub-μA programming current," in *Symp. VLSI Technol.-Dig. Tech. Papers*, 2011, pp. 22–23.

[9] J.-J. Huang, Y.-M. Tseng, W.-C. Luo, C.-W. Hsu, and T.-H. Hou, "One selector-one resistor (1S1R) crossbar array for high-density flexible memory applications," in *IEDM Tech. Dig.*, Dec. 2011, pp. 31.7.1–31.7.4, doi: 10.1109/IEDM.2011.6131653.

[10] L. Zhang *et al.*, "High-drive current (>1 MA/cm^2) and highly nonlinear (>10^3) TiN/amorphous-Silicon/TiN scalable bidirectional selector with excellent reliability and its variability impact on the 1S1R array performance," in *IEDM Tech. Dig.*, San Francisco, CA, USA, Dec. 2014, pp. 6.8.1–6.8.4, doi: 10.1109/IEDM.2014.7047000.

[11] C. Ho *et al.*, "Threshold vacuum switch (TVS) on 3D-stackable and $4F^2$ cross-point bipolar and unipolar resistive random access memory," in *IEDM Tech. Dig.*, San Francisco, CA, USA, Dec. 2012, pp. 2.8.1–2.8.4, doi: 10.1109/IEDM.2012.6478968.

[12] J. Woo *et al.*, "Multi-layer tunnel barrier (Ta_2O_5/TaO_X/TiO_2) engineering for bipolar RRAM selector applications," in *Proc. Symp. VLSI Technol.*, Kyoto, Japan, 2013, pp. T168–T169.

[13] Q. Luo *et al.*, "Cu BEOL compatible selector with high selectivity (>10^7), extremely low off-current (~pA) and high endurance (>10^{10})," in *IEDM Tech. Dig.*, Washington, DC, USA, Dec. 2015, pp. 10.4.1–10.4.4, doi: 10.1109/IEDM.2015.7409669.

[14] A. Goda and K. Parat, "Scaling directions for 2D and 3D NAND cells," in *IEDM Tech. Dig.*, San Francisco, CA, USA, Dec. 2012, pp. 2.1.1–2.1.4, doi: 10.1109/IEDM.2012.6478961.

[15] W. Kim *et al.*, "Multi-layered vertical gate NAND flash overcoming stacking limit for terabit density storage," in *Proc. Symp. VLSI Technol.*, Honolulu, HI, USA, 2009, pp. 188–199.

[16] M. Shibata *et al.*, "A 1.33 Tb 4-bit/cell 3D-flash memory on a 96-word-line-layer technology," in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, San Francisco, CA, USA, Feb. 2019, pp. 210–212, doi: 10.1109/ISSCC.2019.8662443.

[17] S. Lee *et al.*, "A 1 Tb 4 b/cell 64-stacked-WL 3D NAND flash memory with 12 MB/s program throughput," in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, San Francisco, CA, USA, Feb. 2018, pp. 340–342, doi: 10.1109/ISSCC.2018.8310323.

[18] S. Kim *et al.*, "Ultrathin (<10 nm) Nb_2O_5/NbO_2 hybrid memory with both memory and selector characteristics for high density 3D vertically stackable RRAM applications," in *Proc. Symp. VLSI Technol.*, Honolulu, HI, USA, Jun. 2012, pp. 155–156, doi: 10.1109/VLSIT.2012.6242508.

[19] M.-C. Hsieh *et al.*, "Ultra high density 3D via RRAM in pure 28 nm CMOS process," in *IEDM Tech. Dig.*, Washington, DC, USA, Dec. 2013, pp. 10.3.1–10.3.4, doi: 10.1109/IEDM.2013.6724600.

[20] Y.-W. Chin, S.-E. Chen, M.-C. Hsieh, T.-S. Chang, C. J. Lin, and Y.-C. King, "Point twin-bit RRAM in 3D interweaved cross-point array by Cu BEOL process," in *IEDM Tech. Dig.*, San Francisco, CA, USA, Dec. 2014, pp. 6.4.1–6.4.4, doi: 10.1109/IEDM.2014.7046996.

[21] C. Xu, P.-Y. Chen, D. Niu, Y. Zheng, S. Yu, and Y. Xie, "Architecting 3D vertical resistive memory for next-generation storage systems," in *Proc. IEEE/ACM Int. Conf. Comput.-Aided Design (ICCAD)*, San Jose, CA, USA, Nov. 2014, pp. 55–62, doi: 10.1109/ICCAD.2014.7001329.

[22] H.-Y. Chen, S. Yu, B. Gao, P. Huang, J. Kang, and H.-S. P. Wong, "HfO_X based vertical resistive random access memory for cost-effective 3D cross-point architecture without cell selector," in *IEDM Tech. Dig.*, San Francisco, CA, USA, Dec. 2012, pp. 20.7.1–20.7.4, doi: 10.1109/IEDM.2012.6479083.

[23] C.-W. Hsu *et al.*, "Self-rectifying bipolar TaO_X/TiO_2 RRAM with superior endurance over 10^{12} cycles for 3D high-density storage-class memory," in *Proc. Symp. VLSI Technol.*, Kyoto, Japan, 2013, pp. T166–T167.

[24] S. Lee, J. Sohn, Z. Jiang, H.-Y. Chen, and H.-S. P. Wong, "Metal oxide-resistive memory using graphene-edge electrodes," *Nature Commun.*, vol. 6, p. 8407, Sep. 2015, doi: 10.1038/ncomms9407.

[25] K.-S. Li *et al.*, "Utilizing Sub-5 nm sidewall electrode technology for atomic-scale resistive memory fabrication," in *Proc. Symp. VLSI Technol.*, Honolulu, HI, USA, Jun. 2014, pp. 1–2, doi: 10.1109/VLSIT.2014.6894402.

[26] M. Yu *et al.*, "Novel vertical 3D structure of TaO_X-based RRAM with self-localized switching region by sidewall electrode oxidation," *Sci. Rep.*, vol. 6, Feb. 2016, Art. no. 21020, doi: 10.1038/srep21020.

[27] Q. Luo *et al.*, "Super non-linear RRAM with ultra-low power for 3D vertical nano-crossbar arrays," *Nanoscale*, vol. 8, pp. 15629–15636, Aug. 2016, doi: 10.1039/c6nr02029a.

[28] Q. Luo *et al.*, "8-Layers 3D vertical RRAM with excellent scalability towards storage class memory applications," in *IEDM Tech. Dig.*, San Francisco, CA, USA, Dec. 2017, pp. 2.7.1–2.7.4, doi: 10.1109/IEDM.2017.8268315.

[29] Z. Jiang *et al.*, "Selector requirements for tera-bit ultra-high-density 3D vertical RRAM," in *Proc. Symp. VLSI Technol.*, Honolulu, HI, USA, Jun. 2018, pp. 107–108, doi: 10.1109/VLSIT.2018.8510689.

[30] 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. 5th IEEE Int. Memory Workshop (IMW)*, Monterey, CA, USA, May 2013, pp. 155–158, doi: 10.1109/IMW.2013.6582122.

[31] Y. Deng *et al.*, "Design and optimization methodology for 3D RRAM arrays," in *IEDM Tech. Dig.*, Washington, DC, USA, Dec. 2013, pp. 25.7.1–25.7.4, doi: 10.1109/IEDM.2013.6724693.

[32] P.-Y. Chen and S. Yu, "Impact of vertical RRAM device characteristics on 3D cross-point array design," in *Proc. IEEE 6th Int. Memory Workshop (IMW)*, Taipei, Taiwan, May 2014, pp. 1–4, doi: 10.1109/IMW.2014.6849382.

[33] S. Yu *et al.*, "Design guidelines for 3D RRAM cross-point architecture," in *Proc. IEEE Int. Symp. Circuits Syst. (ISCAS)*, Melbourne, VIC, Australia, Jun. 2014, pp. 421–424, doi: 10.1109/ISCAS.2014.6865155.

[34] Q. Hua *et al.*, "Threshold switching selectors: A threshold switching selector based on highly ordered Ag nanodots for X-point memory applications," *Adv. Sci*, vol. 6, no. 10, 2019, Art. no. 1970058, doi: 10.1002/advs.201970058.

[35] Q. Luo *et al.*, "Demonstration of 3D vertical RRAM with ultra low-leakage, high-selectivity and self-compliance memory cells," in *IEDM Tech. Dig.*, Washington, DC, USA, Dec. 2015, pp. 10.2.1–10.2.4, doi: 10.1109/IEDM.2015.7409667.

[36] X. Xu *et al.*, "Fully CMOS compatible 3D vertical RRAM with self-aligned self-selective cell enabling sub-5nm scaling," in *Proc. IEEE Symp. VLSI Technol.*, Honolulu, HI, USA, Jun. 2016, pp. 1–2, doi: 10.1109/VLSIT.2016.7573388.

[37] R. Yamashita *et al.*, "A 512 Gb 3b/cell flash memory on 64-word-line-layer BiCS technology," in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, San Francisco, CA, USA, Feb. 2017, pp. 196–197, doi: 10.1109/ISSCC.2017.7870328.

[38] C. Kim *et al.*, "A 512 Gb 3b/cell 64-stacked WL 3D V-NAND flash memory," in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, San Francisco, CA, USA, Feb. 2017, pp. 202–203, doi: 10.1109/ISSCC.2017.7870331.

[39] T. Tanaka *et al.*, "A 768Gb 3b/cell 3D-floating-gate NAND flash memory," in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, San Francisco, CA, USA, Jan./Feb. 2016, pp. 142–144, doi: 10.1109/ISSCC.2016.7417947.

[40] *ASU ASAP7 PDK*. Accessed: Aug. 24, 2017. [Online]. Available: http://asap.asu.edu/asap/

[41] L. Zhang, "Study of the selector element for resistive memory," Ph. D. dissertation, Dept. Elect. Eng., KU Leuven, Leuven, Belgium, 2015.

[42] Z. Jiang *et al.*, "Next-generation ultrahigh-density 3-D vertical resistive switching memory (VRSM)—Part II: Design guidelines for device, array, and architecture," *IEEE Trans. Electron Devices*, to be published, doi: 10.1109/TED.2019.2950595.

[43] *Sentaurus Device Simulator User Guide*, document Version J-2014.09, Mountain View, CA, USA, Sep. 2014.

[44] L. Zhao *et al.*, "Ultrathin (~2nm) HfO_X as the fundamental resistive switching element: Thickness scaling limit, stack engineering and 3D integration," in *IEDM Tech. Dig.*, San Francisco, CA, USA, 2014, pp. 6.6.1–6.6.4, doi: 10.1109/IEDM.2014.7046998.

[45] W. Banerjee *et al.*, "Occurrence of resistive switching and threshold switching in atomic layer deposited ultrathin (2 nm) aluminium oxide crossbar resistive random access memory," *IEEE Electron Device Lett.*, vol. 36, no. 4, pp. 333–335, Apr. 2015, doi: 10.1109/LED.2015.2407361.

[46] B. Govoreanu *et al.*, "Performance and reliability of ultra-thin HfO_2-based RRAM (UTO-RRAM)," in *Proc. 5th IEEE Int. Memory Workshop (IMW)*, Monterey, CA, USA, May 2013, pp. 48–51, doi: 10.1109/IMW.2013.6582095.

[47] P. Huang *et al.*, "Self-selection RRAM cell with sub-μA switching current and robust reliability fabricated by high-K/metal gate CMOS compatible technology," *IEEE Trans. Electron Devices*, vol. 63, no. 11, pp. 4295–4301, Nov. 2016, doi: 10.1109/TED.2016.2612824.

[48] B. Gao *et al.*, "A novel self-selection bipolar RRAM cell with ultra-low operation currents for cross-point application," in *Proc. IEEE 11th Int. Conf. Solid-State Integr. Circuit Technol.*, Xi'an, China, Oct./Nov. 2012, pp. 1–3, doi: 10.1109/ICSICT.2012.6467645.

[49] C.-W. Hsu, T.-H. Hou, M.-C. Chen, I.-T. Wang, and C.-L. Lo, "Bipolar Ni/TiO_2/HfO_2/Ni RRAM with multilevel states and self-rectifying characteristics," *IEEE Electron Device Lett.*, vol. 34, no. 7, pp. 885–887, Jul. 2013, doi: 10.1109/LED.2013.2264823.