Design Trad-Offs for High Density Cross-Point Resistive Memory

***Abstract*—With conventional memory technologies approaching their scaling limit, emerging non-volatile memory technologies have attracted considerable attention because of their non-volatility, high access speed, low power consumption, and good scalability. Resistive RAM (ReRAM), with its simple structure, small cell size ($4F^2$), and support for 3D stacking, has been a leading candidate among emerging technologies. A key advantage of ReRAM comes from its non-linear nature, which enables us to build a cross-point RAM array without having a dedicated access transistor in each cell. While cross-point design is effective in improving memory density, it has inherent disadvantages which introduce extra design challenges. Based on the circuit characteristics of the cross-point array, we propose a mathematical model to perform a comprehensive analysis of issues of reliability, energy consumption and the area overhead. In addition to the cell-level analysis, different programming schemes are also discussed in detail. The proposed model enables designers to identify the most energy/area efficient ReRAM organization that meets specific design goals in the early design stage.**

I. Introduction

The scaling of traditional memory technologies, such as DRAM and FLASH, is approaching its physical limit. In the past few years, emerging non-volatile technologies (NVM), such as Phase Change RAM (PCRAM), Spin-transfer-torque RAM (STT-RAM), and Resistive RAM (ReRAM) have been widely studied as potential candidates for the next generation memory technologies to meet the need of higher density, faster access time, and lower power consumption. Among all of these emerging memory technologies, ReRAM has many unique characteristics, including simple structure, non-linearity and high resistance ratio, making itself one of the most promising technologies. Researchers have shown that the state-of-the-art single-level-cell ReRAM can achieve 7.2ns random access time for both read and write operations with a resistance ratio larger than 100 [1]. Also, HP labs and Hynix have already announced plans to commercialize the memristor-based ReRAM and predicted that ReRAM could eventually replace traditional memory technologies [2].

Unlike other non-volatile memory technologies, ReRAM can be implemented in a cross-point style structure without any access device. Specifically, in a nano cross-point array, each bistable ReRAM cell is sandwiched by two orthogonal nanowires. Thus the area occupied by each cell is literally the area underneath the intersection of wires, which is $4F^2$ per bit. However, the simplicity of access device free, cross-point structure introduces challenges to the peripheral circuit and memory organization design.

While there have been prior studies on cross-point ReRAM array [3]–[6], they do not consider the effect of voltage drivers and programming methods to the array. In addition, detailed area and energy analysis is also absent. In this work, we address the design challenges of cross-point structure based ReRAM. We build an accurate mathematical model to evaluate memory reliability, energy consumption, and area overhead for different designs and cell parameters. The advantages of non-linearity and write current scaling are all discussed in detail. Our study allows for exploring the most energy/area efficient ReRAM design with different design constraints and cell parameters at the very beginning of the design stage. On the other hand, system designers can also leverage the proposed model to provide valuable feedback to device researchers who will in turn adjust ReRAM cell design. We believe that this kind of collaboration will be very helpful to shorten time-to-market of ReRAM memory.

The rest of this paper is organized as follows. In Section II, the preliminaries of ReRAM technology and cross-point architecture are introduced. Section III discusses the mathematical model we propose for crossbar structure ReRAM and the edge conditions for different write and read schemes. Section IV analyzes different design constraints of write and read operations on the cross-point based ReRAM array. The energy consumption and area overheads are also analyzed in this section. Then in Section V, the effect of non-linearity and write current on the design constraints are evaluated. Finally, the conclusion is presented in Section VI.

II. Preliminaries

This section provides background of ReRAM and cross-point architecture, and discusses their advantages and limitations.

A. Background of ReRAM technology

As implied by its name, a ReRAM cell uses its resistance to represent the stored information. A ReRAM cell can be switched between high resistance state (HRS) and low resistance state (LRS) by applying an external voltage across the cell. In general, a ReRAM cell is built on a Metal-Insulator-Metal (MIM) structure. The resistance switching behaviors have been observed in many MIM nanodevices with different metal oxide materials. For example, a particular TiO_2 based MIM structure ReRAM, named 'memristor', was developed by HP Labs in 2008 [7]. The proposed memristor-based ReRAM is considered as the first experimental realization and a theoretical model of the fourth fundamental circuit element, which is predicted by Chua [8] about 40 years ago. The memristor-based ReRAM has very small cell size with an access time of less than 50ns. Another H_fO_2-based bipolar ReRAM prototype was fabricated by ITRI this year with an access time as low as 7.2ns [1].

Although there are several variants of ReRAM cells, all of them can be classified into two broad categories: unipolar ReRAM and bipolar ReRAM. In an unipolar cell, the resistance switching behaviors do not depend on the polarity of the voltage input across the cell and only relate to magnitude and duration of the voltage input. On the other hand, in a bipolar cell, the voltage polarity for ON-to-OFF switching (RESET operation) is different from OFF-to-ON switching (SET operation). The need for different pulse widths for SET and RESET in unipolar ReRAM means that its write latency will be determined by the longest pulse. Moreover, the control of SET, RESET, and read operations without any disturbance is another crucial design challenge, especially in high speed ReRAM design. For these reasons, most high performance ReRAM studies are dominated by bipolar ReRAM [1], [9], [10]. In this study, we perform a detailed analysis of the design challenges of bipolar ReRAM cross-point arrays.

B. Cross-Point Architecture

There are two possible memory structures for a bipolar ReRAM implementation: a traditional MOSFET-accessed structure and a cross-point structure. In the MOSFET-accessed memory array, a MOSFET is used as an access device for each memory cell. As the size of a MOSFET access device is typically much larger than the size of a ReRAM cell, the total area of memory array is primarily dominated by MOSFETs rather than ReRAM cells. Also, in order to prove enough driven current, lager than minimum-sized transistor should be used for writes. Hence, ReRAM's area advantage gets lost

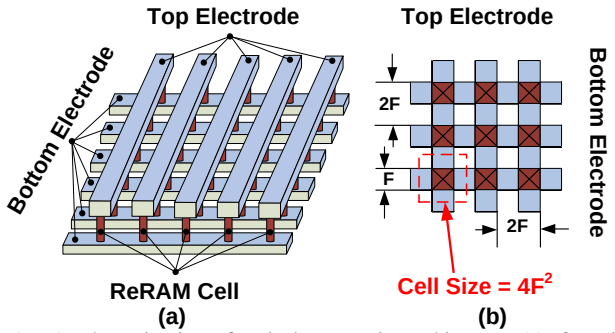

Fig. 1. A schematic view of typical cross-point architecture. (a): Overview of the cross-point architecture. (b): The layout of the cell size of $4F^2$.

because of the access device. In contrast, the cross-point structure is more area-efficient for the ReRAM memory array [11]. A schematic view of a typical cross-point memory array is shown in Figure 1(a). In a cross-point array, ReRAM is sandwiched between wordlines and bitlines. Figure 1(b) shows the feasibility of $4F^2$ cells, the theoretical minimum size for a single layer single level memory cell. The memory density can further be improved by using a multi-layer multi-level cross-point ReRAM array [**?**] [12].

In a cross-point structure, the write operation can either write one bit per access or several bits attached to a wordline at the same time. Although the second scheme has higher bandwidth, it requires a two-step write operation to prevent unintentional writing [11], significantly increasing the write latency. While writing to a cross-point array, the unselected wordlines and bitlines can either be left floating or half biased. On the other hand, while reading a cell, the selected wordline should be biased at read voltage and all the other wordlines and bitlines in the array are shunted to ground. The current in each bitline is then sensed and compared to a reference current to determine cell content. However, due to the sneak current existing at the cross-point array, the current in bitline also varies depending upon the values stored in unselected cells. This phenomenon of read disturbance restricts the size of a cross-point array, since sneak current increases with the number of cells attached to wordlines and bitlines. In order to mitigate this problem, a two-step read operation was proposed: in the first step, the background current of the cross-point array will be sensed. Then the total current, comprising of both the background current and current at the selected cell, will be read out. The state of the selected cell can then be determined by computing the difference between the total current and background current. For this scheme to work, the cross-point array should be sized such that the difference between background current and current through the selected cell should be large enough for reliable sensing.

III. Modeling of the Cross-Point Memory

The basic circuit model of an M by N cross-point ReRAM array is shown in Figure 2. The model is built upon Kirchhoff's Current Law (KCL) and its validity can be guaranteed by deductions from the basic circuit theory. The horizontal lines are wordlines and vertical lines represent bitlines. The ReRAM cells are located at each cross-point of wordline and bitlines. The resistance of the ReRAM cell at the cross-point of i^{th} wordline and j^{th} bitline is represented by $R_{i,j}$. We assume the resistance of the wire connecting two cross-points to be R_{line}. The input resistance of each wordline and bitline is R_v and the resistance of sense amplifier is R_s. In order to set up the KCL equations, the voltage at each cross-point is indicated as $V_{i,j}$ for wordline and $V'_{i,j}$ for bitline. A detailed cross-point is also shown in Figure 2(b). The input voltage for the i^{th} wordline is V_{Wi} and the i^{th} bitline is V_{Bi}. In the case where a wordline takes input from both the sides, the voltage at the other end of the i^{th} wordline is represented as V'_{Wi}.

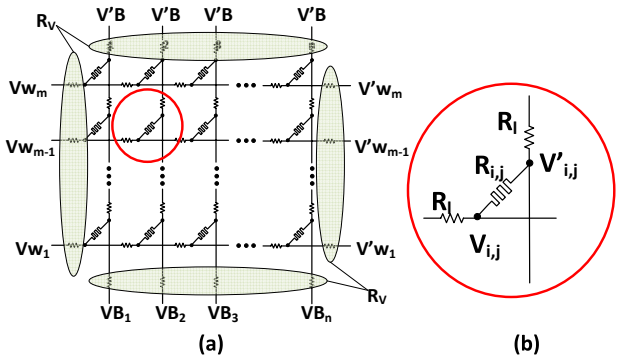

Fig. 2. The basic model of typical cross-point array.

Based on this model, the current equations for each cross-point can be set following KCL: $\Sigma_{I=1}^{k} I_k = 0$. All of the cross-points have similar structure with no more than three current branches and therefore it is very easy to set up the KCL equations for each cross-point. However, we should treat the cross-points at the edges of the array specifically because KCL equations for these cross-points vary with different write/read schemes. For example, the unselected wordline for write operation can be either half biased or left floating. Thus, the edge conditions should be adjusted according to each write/read scheme. In particular, all of the cross-points in an array can be classified into three major categories: *normal point*, *activated point* and *floating point*.

The normal points are located inside the memory array. In other words, for all of the nodes with $1 < i < m$ and $1 < j < n$, the KCL equations take the form of

$$R_l^{-1}V_{i,j-1} - (2R_l^{-1} + R_{i,j}^{-1})V_{i,j} + R_l^{-1}V_{i,j+1} + R_{i,j}^{-1}V'_{i,j} = 0, \quad (1)$$

for the node at wordline layer and

$$R_l^{-1}V'_{i-1,j} - (2R_l^{-1} + R_{i,j}^{-1})V'_{i,j} + R_l^{-1}V'_{i+1,j} + R_{i,j}^{-1}V_{i,j} = 0, \quad (2)$$

for the node at bitline layer.

The activated point and floating point represent the nodes at the edge of cross-point array with different conditions: an edge point, which is directly connected to the voltage input or to the ground, can be considered as an activated point. Otherwise, it is a floating point. For example, consider the point located at the intersection of i^{th} wordline and 1^{st} bitline. If the i^{th} wordline is activated by an input voltage of V_{Wi}, this cross-point is an activated point, and the KCL equation for this point is:

$$-(R_v^{-1}+R_l^{-1}+R_{i,1}^{-1})V_{i,1}+R_l^{-1}V_{i,2}+R_{i,1}^{-1}V'_{i,1} = -R_v^{-1}V_{Wi}. \quad (3)$$

Otherwise, it is floating and its KCL equation is

$$-(R_l^{-1} + R_{i,1}^{-1})V_{i,1} + R_l^{-1}V_{i,2} + R_{i,1}^{-1}V'_{i,1} = 0. \quad (4)$$

For clarity, a $2mn \times 1$ vector V is defined to represent all of the variables in the KCL equations:

$$V = [V_1{}^T, V_2{}^T ... V_m{}^T, V_1'^T, V_2'^T ... V_m'^T]^T, \quad (5)$$

where,

$$V_i = [V_{i,1}, V_{i,2} ... V_{i,n}]^T, \quad V'_i = [V'_{i,1}, V'_{i,2} ... V'_{i,n}]^T, \quad (6)$$

for $i = 1, 2...m$. Then all of the KCL equations can be considered as a system of linear equations, which has the form

$$A \cdot V = C. \quad (7)$$

A is a $2mn \times 2mn$ coefficient matrix, which is determined by Equations(1)-(4). C is a $2mn \times 1$ vector, containing the constant terms of these equations. As shown, all of the KCL equations have

simple structure and are similar to each other. Therefore, the linear equation system has a relatively fixed format and simple structure, making it easy to establish and adjust the coefficients and constants according to different design schemes. Besides, due to the simplicity of the KCL equation, A is populated primarily with zeros and can be saved as a sparse matrix, which will further reduce the storage cost during the computation.

To validate our analytical model, we compared the results with the HSPICE simulations using simple a resistor model in cross-point memory arrays. DC analysis was performed by HSPICE which solved the voltage of every node in the array. The results of eight cross-point arrays with different array size and specific data pattern are shown in Figure ??, the voltage drop on the selected cell derived by our analytical model are consistent with the HSPICE simulation results.

Thus, with parameters such as the resistance of ReRAM cells, the resistance of interconnect wires, program voltages, and write/read schemes, voltages at various cross points can be obtained by solving the system of linear equations. With detailed voltage values, $V_{2mn\times 1}$, we can analyze the array at a fine granularity. These values are also critical to evaluate reliability, energy consumption, driven current density, and area overheads of a cross-point array.

IV. Analysis of Design Constraints - A Case Study

In this section, we study the effect of various schemes on cross-point size and reliability in detail by using our mathematical model. The constraints on array size, energy consumption and area overhead are analyzed in the worst cases scenario. The results of this study will be a useful guide in designing a cross-point array.

A. Overview

In order to write or read a cross-point array, proper voltages should be applied across the ReRAM cell. Although the goal of a read operation is different from a write operations, both of them are realized by fully biasing the selected wordlines/bitlines and floating (or half biasing) unselected wordlines/bitlines. Thus, the coefficient matrix A and the constant vector C are very similar for both. In addition, their energy consumption and area overhead will also have a similar trend. Therefore, in this section, we first study the write operation comprehensively. After that, for read operation, we mainly focus on the read margin analysis since it is unique to reads.

Table I shows the circuit parameters of our baseline 50nm design. The data is derived from the recently published studies on ReRAM [**?**], [11]. We study reliability, energy consumption, and area overheads for four different write schemes, and discuss the sensitivities of these schemes to the data pattern of HRS and LRS ReRAM cells and cell non-linearity.

TABLE I
Parameters of the baseline Cross-Point Array

Metric	Description	Typical Values (Range)
S_{cell}	Cell Size	$4F^2$
R_l	Interconnection Resistance	0.65Ω
V_{RESET}	Threshold voltage for RESET	$2.0V$
V_{SET}	Threshold voltage for SET	$-2.0V$
V_{READ}	Read Voltage of Cell	$0.5V$
I_{on}	Write Current for LRS Cell	$40uA$ $(20 \sim 200uA)$
$V_W(R)$	Wordline Voltage during Read	$0.4V$
$V_W(W)$	Wordline Voltage during Write	$\pm 2V$
$V_W(H)$	Half Selected wordline Voltage	$1V$
$V_B(R)$	Bitline Voltage during Read	$0V$
$V_B(W)$	Bitline Voltage during Write	$0V$
$V_B(H)$	Half Selected bitline Voltage	$1V$
K_r	Nonlinearity of ReRAM Cell	5 $(2 \sim 40)$
M, N	Number of wordlines/bitlines	512 $(8 \sim 1024)$

B. Write Operation

To write a ReRAM cell, an external voltage is applied across the cell for a certain duration. Intuitively, there are four possible schemes for the write operation:

1) According to the location of a selected cell, activate one wordline and one bitline and leave all of other lines floating (FWFB shemes).
2) Activate the selected wordline and bitline. Leave all the unselected wordlines floating and half bias the unselected bitlines (FWHB shemes).
3) In contrast to the scheme 2), activate the selected wordline and bitline. Leave all the unselected bitlines floating and half bias the unselected wold lines (HWFB shemes).
4) Activate the selected wordline and bitline. Then half bias all other wordlines and bitlines (HWHB shemes).

However,the FWFB scheme has inherent problem that may result in severe write disturbance [**?**]. Therefore, in the following discussion, we only compare the results of FWHB, HWFB and HWHB schemes. For each of these three schemes, we can either write several cells at one wordline at the same time or only write one bit per access and distribute the write operation to several arrays. In the following discussion, we start from one bit per access write operation, then the results of one wordline per access method are discussed.

Reliable Write Operation.

Write reliability is a serious concern in cross-point arrays. In an ideal condition, the resistance of wires and the sneak currents in unselected cells are negligible. In such a scenario, all the write schemes discussed above will make sure that the write voltage $V_W(W) - V_B(W)$ is fully applied across the specified cell. However, in reality, both wire resistance and sneak current are non-trivial. Hence, the operation of cross-point array will vary based on the data pattern stored in ReRAM cells. A write is considered reliable if it modifies the content of the selected cells to the new value without disturbing other unselected cells. There are two potential problems with writes: *write failure*, an unsuccessful write on selected cell, and *write disturbance*, an undesirable write on unselected cell. It is necessary to ensure that a write scheme guarantees reliable operation even in the worst case (w.r.t the location of cells to written and the data pattern stored in the cross-point array). Otherwise, after several unreliable write operations, the data stored in the cross-point array will become unpredictable.

Write failure typically results from the voltage drop at the interconnect wires along the wordline and bitline. It has been shown that [5], for one bit per access write operation, the worst case voltage drop occurs when writing the cell at the cross point of the M^{th} wordline and the N^{th} bitline with all of the cells in the array are in LRS. In order to avoid the write failure and successfully program the selected ReRAM cell, the driven voltage should be boosted to a higher level, making sure that the voltage across the cell exceeds the threshold voltage even at the worst case. Figure 3 shows the lower bound of the driven voltage for different sizes of cross-point array. The minimum wordline/bitline voltage increases from 2.01 V for a 8×8 array to 4.2 V for a 1024×1024 cross-point array. In addition, for a memory capability, the cross-point array can be organized with different number of wordlines and bitlines. For example, a 256K bits cross-point array can be implemented either by a 512×512 array or by a 64×4096 array. In the latter case, the voltage drops along the wordline will be much more serious than along the bitline. Figure 4 examines the voltage requirement for different array organizations with different write schemes. The result shows that from a reliability point of view, a cross-point array with same numbers of wordlines and

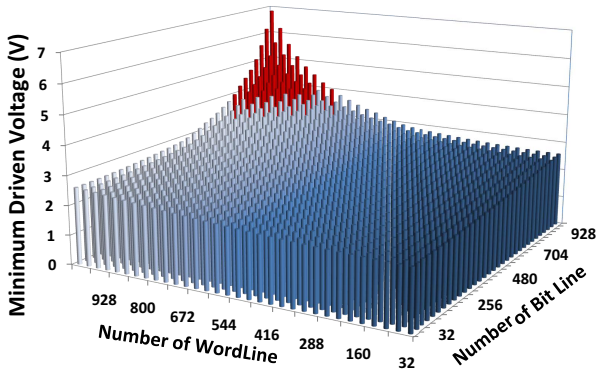

Fig. 3. Write voltage requirement (Threshold voltage = 2V).

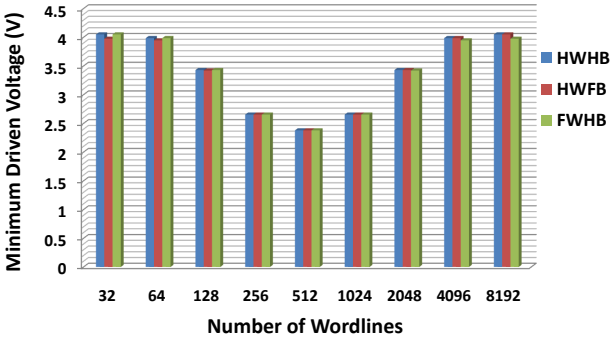

Fig. 4. Write voltage requirement with different memory shape. (Array capacity = 256Kbits, Threshold voltage = 2V.)

bitlines is the best choice. Furthermore, we also notice that when the array has the same number of wordlines and bitlines, FWFB, HWFB and FWHB schemes have the same minimum driven voltage.

However, boosting the driven voltage also introduces other potential problems for the array. In particular, increasing the driven voltage also increases the voltage applied at unselected cells. Therefore, a write disturbance may occur when the voltage applied at an unselected cell exceeds the threshold voltage for SET or RESET operation. Specifically, the maximum voltage applied at unselect cells is exactlly the same as half of the driven voltage. Thus, only the array with driven voltage less than 4V are allowable. Otherwise, the array is unreliable because it can not avoid write failure and write disturbance at the same time. The unreliable array size of are denoted as red bars in Figure 3. The array size limitation provied by Figure 3 is a hard constraint on array size, and all of the following energy and area tradeoffs should be bounded by this constraint.

Energy Consumption of Write Operation.

The energy consumption of a write operation includes: the energy consumed to change the state of the selected cell (denoted as E_{select}), the undesired energy wasted at the half selected cells ($E_{halfselect}$) and unselected cells ($E_{unselect}$), as well as the energy consumed by the interconnect lines (E_{line}). Figure 5 shows the decomposed energy consumption for the cross-point array. Note that, the E_{line} and $E_{halfselect}$ take a large amount of the total energy consumption. Also, this part of energy wasted during the write operation takes greater part of the total energy for larger array sizes. For example, the undesired energy consumption for writing a 512×512 array is more than 1000 times larger than the 8×8 array. We also notice that, since the impact of sneak paths for floating schemes (FWHB and HWFB) is more serious, the energy consumed at unselected cells for floating schemes is larger than the half-biased scheme. Due to this reason, the total energy consumptions for FWHB and HWFB

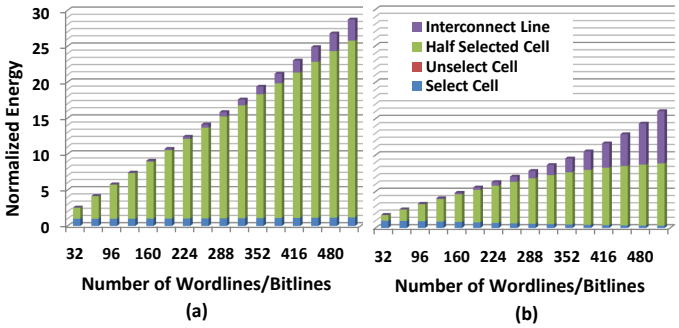

Fig. 5. The normalized energy consumption. (a): One bit writing; (b): Multi-bit writing.

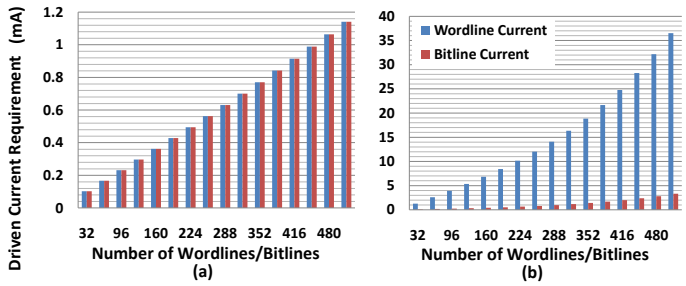

Fig. 6. The driven current requirements for wordlines and bitlines. (a) One bit writing; (B) Multi-bit writing.

schemes are at least 10% larger than that of HWHB scheme.

Write Current and Area Overhead of Write Operation.

The write operation for a $M\times N$ array requires M wordline voltage drivers and N bitline multiplexers. The drivers and multiplexers should be sized such that they can provide the worst-case current of wordline current and bitline current. The transistor sizing of the wordline/bitline circuitry are achieved using HSPICE simulations. We further calculate the area overhead for the drivers and multiplexers by referring to the CACTI area model. Figure 6(a) shows the maximum write current with different ReRAM array sizes. As we can see, the current requirement increases slightly faster than linearly as array size increases. This is primarily because we have to increase the write voltage to compensate the voltage drop in large array size. Figure 6(b) demonstrates the area overhead for the wordline and bitline circuitry. It is indicated that wordline drivers occupies comparable or even larger area than the memory cells, degrading the effective cell size significantly. There are two possible reasons: (1) wordline drivers are usually implemented as tri-gates and larger than bitline multiplexers that are essentially pass transistors; (2) wordline drivers have to provide more than one set of write current in multi-bits write operation, as discussed in the following section.

Discussion on Multi-Bits Write Operation.

So far, we have only discussed the write operation with one bit per access. In this section, we consider the difference between one bit per access and one wordline per access write operations. Firstly, writing a wordline at a time will worsen the voltage drop along the wordline. Therefore the simulation results show that the reliable size of the cross-point array will be further reduced. The maximum array size reduces from 116×116 to 100×100 for HWFB and HWFB schemes.

In order to fairly compare the energy consumption, we compare the energy-per-bit instead of the total energy. For example, in order to write a wordline with size of 128, the energy-per-bit can be calculated as: $E_{ave} = E_{total}/128/2$. Figure 8 shows the energy-per-bit of

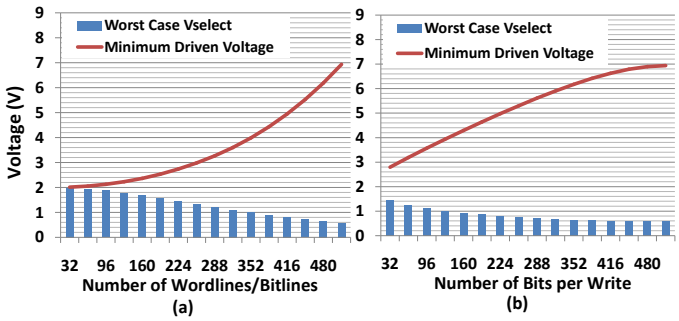

Fig. 7. Worst case voltage and write voltage requirment for multi-bit writing. (a) One wordline wrting with different array size; (b) 512x512 array with various number of bits per writing.

the multi-bit write operation. The energy shown in this figure is normalized to the same unit as Figure 5 for easier comparison. The results show that for large cross point array sizes, the multi-bit write operation is much more energy efficient. This is because the energy wasted at the unselected and half-selected cells are shared by multiple bits and the average energy for one each bit is therefore reduced. However, although the multi-bit write operation has the advantage of lower energy consumption, the maximum current requirement for each wordline also increases. As shown in Figure 6(b), the maximum driven current for each bitline is almost the same as when writing one bit, however the drive capability for each word line is almost doubled for multi-bit writing. Since the area of the voltage driver increases proportionally with its drive capability, the area overhead for multi-bit writing is about 50% larger than for one bit writing.

C. Read Operation

In this section we applied the similar sensing scheme as [5] and [4]. In order to read cell $R_{i,j}$, the i^{th} wordline is biased at V_{READ} and all of the other wordlines and bitlines are grounded. Then the state of the selected cell is read out by measuring the voltage across R_s. The energy consumption for read operation can be analyzed by the same way as that of the write operation. Since the read voltage is much smaller than write voltage, the read energy is expected at least one order smaller than write operation. Considerable sensing margin is achieved by implementing a current-to-voltage converter and sensing the voltage signal using traditional or emerging sense amplifier design. The input resistance of the current-to-voltage converter is extracted from HSPICE simulation results. Read sensing margin is defined as $\Delta V = \Delta I \times R_{converter}$ where $R_{converter}$ is input resistance of the converter.

Additionally, since the read voltage/current is much lower than the write, we believe that the voltage drivers can always provide enough current for the read operation if they meet the current requirement for write operation. Therefore, we can conclude that the area overhead of voltage drivers is determined by the write current. However, the reliability of read operation is different from the write operation. The read reliability is determined by the voltage swing for reading HRS and LRS cells. Figure **??** (a) shows the voltage swing with different array sizes. Specifically, larger cross-point array have more sneak paths, making the output voltage very sensitive to the data pattern of unselected cell. Therefore, the sense marging decreases with the increase of array size. In order to improve the reliability of the read operation, a two-step sensing scheme can be applied, which senses the current of an unselected cell first, then the overall current is sensed, and after that the current difference is converted to the output voltage. The voltage swing of this two-step sensing scheme is shown in Figure **??** (b). By using this two-step sensing schemes, the voltage swing for a given array size and non-linearity coefficient is doubled.

V. Non-linearity and Write Current Scaling

(**Figure and discussion in this section should be modified after we got the results**) One of the most distinct features of ReRAM is its non-linearity. For example, the non-linearity of memristor-based ReRAM is observed at LRS when the resistance of the memristor cell is not constant but varies with the applied voltage. The non-linearity coefficient is defined as: $K_r(p,V) = p \times R(V/p)/R(V)$, where $R(V/p)$ and $R(V)$ are the equivalent resistance of the memristor biased at V/p and V [11]. Normally, the $K_r(p,V)$ value for memristor-based ReRAM is larger than 20, meaning that the resistance of a half-biased cell is at least 10 times larger than a full-biased cell. Clearly, the ReRAM cell with a larger non-linearity coefficient results in a better memory cell since the current in the sneak path will be significantly reduced. In addition, the increased resistance at half-selected and unselected cells can also mitigate the voltage drop along the activated wordline and bitline. Besides, we found that the cross-point array design can also benefit from the scaling of the write current. Figure 9 shows the influence of different non-linearity coefficients and write current on the array size requirements for one bit HWHB writing scheme. In this figure, the maximum array size increases from 112×112 to 340×340 when the non-linearity coefficient K_r increases from 1 to 10. Similarly, the non-linearity can also increase the maximum array size for other write schemes.

Moreover, the increase of non-linearity or scaling of write current can also reduce the energy consumption and area overheads of the cross-point array. For example, consider a 128×128 array. As shown in Figure 10, the energy consumption for the write operation decreases dramatically with the increase of non-linearity coefficient K_r. As K_r increases from 1 to 40, the write energy is reduced by 98.3%. The driven current requirement is shown in Figure 11(a), and the corresponding area overheads of the voltage drivers are compared to the array size at Figure 11(b). The baseline design is unacceptable because the area of voltage drivers is about 11.6 times larger than the area of the cross-point array. In this case, the area efficiency of ReRAM's $4F^2$ cell size will be offset by the extremely huge area overhead of the voltage drivers. However, with the increase of non-linearity, the area of voltage drivers becomes comparable to the array area. Therefore, we can conclude that, the ReRAM cells with a small non-linearity coefficient are not suitable for the cross-point structure based memory array. Next, we study the area overhead of multi-bit write. Figure 12 shows the normalized areas of the voltage drivers for one bit and multi-bit write operations. As mentioned, multi-bit write operations require larger driven current. Therefore, the area of voltage drivers for multi-bit write operations are much larger than that for one bit write operations. Finally, normalized areas of the one

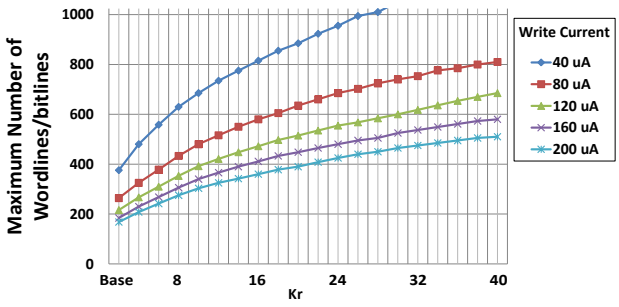

Fig. 8. The maximum array size with different non-linearity coefficient.

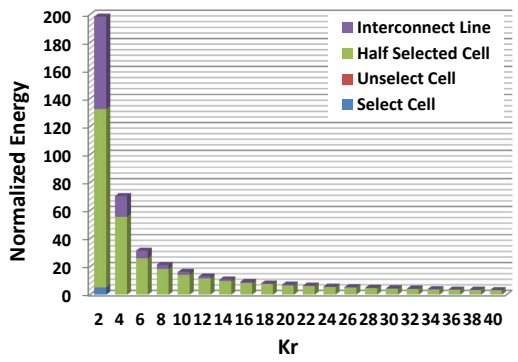

Fig. 9. The normalized energy consumption with non-linear ReRAM cells.

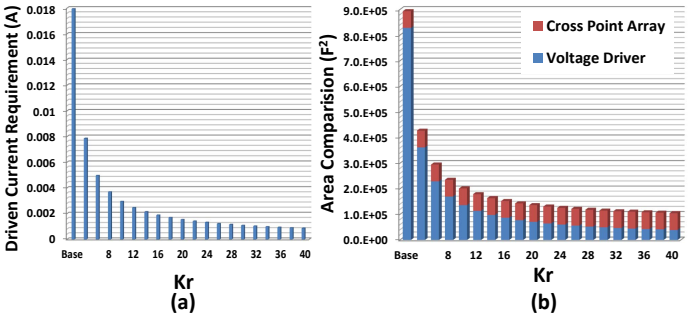

Fig. 10. The driven current requirements and area overheads with different non-linearity coefficients

bit and multi-bit write operations have opposite trends as the array size increases. Normalized area for one bit write operation increases with the array size. On the contrary, normalized area for multi-bit write decreases as the array size increase.

Different from write operation, the read operation can not benefit from the increasing of non-linearity or the scaling of write current. Figure 13 (a) shows the voltage swing with different K_r values and write current. Large non-linearity and small write current are harmful to the voltage swing: **add discussion here after we got the data**

VI. Conclusion

ReRAM is a promising candidate for next-generation non-volatile memory technology. The area efficient cross-point structure is the most attractive memory organization for ReRAM memory design. However, problems inherent in the cross-point structure, such as the existence of sneak current and voltage drops along the nanowires introduce challenges to the design of reliable ReRAM cross-point array. In this paper, we first establish a mathematical model for

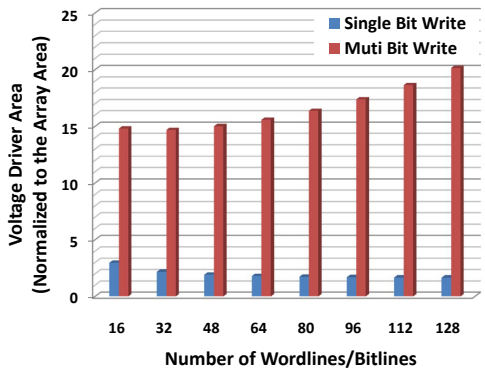

Fig. 11. The normalized area overhead of voltage drivers ($K_r = 20$, the areas are normalized to the area of cross-point array).

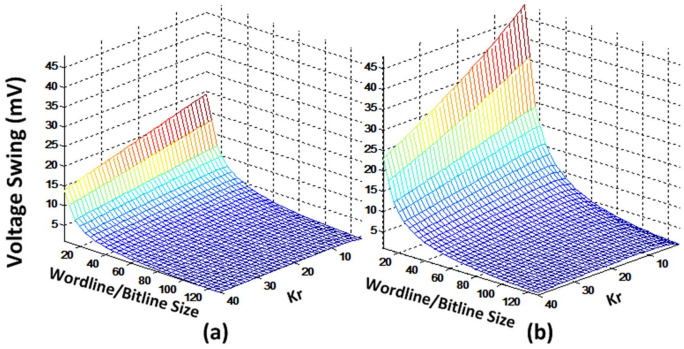

Fig. 12. Relationships among the voltage swing, array size and non-linearity. (a) Normal sensing scheme; (b) Two-step sensing scheme

cross-point arrays. We show that the proposed model has a simple structure and is flexible to evaluate different write/read schemes. By using this model, we study in detail how reliability affects the array organization, size, energy consumption, and area overheads in designing cross-point arrays. The simulation results show that, the multi-bit write operation is more energy efficient than one bit write operation and therefore is more suitable for energy-constrained design. However, from an area-constrained design, one bit write operation is better. Also, we point out that both of the increase of non-linearity and scaling of write current of the ReRAM cell can reduce the energy consumption and area overhead significatively, and it is favorable for large, energy efficient ReRAM design.

References

[1] S. S. Sheu *et al.*, "A 4mb embedded slc resistive-ram macro with 7.2ns read-write random-access time and 160ns mlc-access capability," in *Solid-State Circuits Conference Digest of Technical Papers (ISSCC), 2011 IEEE International*, Feb 2011.

[2] "http://www.hpl.hp.com/news/2010/jul-sep/memristorhynix.html."

[3] M. Ziegler and M. Stan, "Design and analysis of crossbar circuits for molecular nanoelectronics," in *Nanotechnology, 2002. IEEE-NANO 2002. Proceedings of the 2002 2nd IEEE Conference on*, 2002, pp. 323 – 327.

[4] A. Flocke *et al.*, "A fundamental analysis of nano-crossbars with non-linear switching materials and its impact on tio2 as a resistive layer," in *Nanotechnology, 2008. NANO '08. 8th IEEE Conference on*, Aug 2008, pp. 319 –322.

[5] J. Liang and H.-S. Wong, "Cross-point memory array without cell selectors -device characteristics and data storage pattern dependencies," *Electron Devices, IEEE Transactions on*, vol. 57, no. 10, pp. 2531 –2538, Oct 2010.

[6] M. Ziegler and M. Stan, "Cmos/nano co-design for crossbar-based molecular electronic systems," in *Nanotechnology, IEEE Transactions on*, vol. 2, no. 4, Dec 2003, pp. 217 – 230.

[7] D. B. Strukov and et al, "The missing memristor found," in *Nature*, 2008.

[8] L. Chua, "Memristor-the missing circuit element," *IEEE Transactions on Circuit Theory*, no. 5, Sep 1971.

[9] M. Kim *et al.*, "Low power operating bipolar tmo reram for sub 10 nm era," in *Electron Devices Meeting (IEDM), 2010 IEEE International*, Dec 2010.

[10] W. Otsuka *et al.*, "A 4mb conductive-bridge resistive memory with 2.3gb/s read-throughput and 216mb/s program-throughput," in *Solid-State Circuits Conference Digest of Technical Papers (ISSCC), 2011 IEEE International*, Feb 2011.

[11] C. Xu *et al.*, "Design implications of memristor-based rram cross-point structures," in *Proceedings of Design Automation Test in Europe Conference Exhibition 2011*, 2011.

[12] M.-J. Lee *et al.*, "Stack friendly all-oxide 3d rram using gainzno peripheral tft realized over glass substrates," in *Electron Devices Meeting, 2008. IEDM 2008. IEEE International*, Dec 2008, pp. 1 –4.