Design Framework for Reliable, Energy Efficient Cross-Point based Resistive Memory

***Abstract*—Since the conventional memory technologies approaching their scaling limit, the non-volatile memory technologies, such as Phase Change RAM (PCRAM), Magnetoresistive RAM (STT-RAM) and Resistive RAM (ReRAM) have attracted considerable attention because their non-volatility, high access speed, low power consumption and good scalability. Among these emerging memory technologies, the ReRAM has shown great potentials as one of the most promising candidates for future universal memory, due to its simple structure, small cell size and potential for 3D stacking. Besides, the unique non-linearity of ReRAM provides the possibility to build a cross-point structure based ReRAM without CMOS access device, with the smallest cell size of $4F^2$. However, the cross-point structure also suffers from its inherent disadvantages and brings in extra design challenges. In this work, the design challenges of cross-point structure based ReRAM are analyzed. Based on the circuit characteristics of the cross-point array, a precise mathematical model is built to perform a comprehensive analysis on the issues of reliability, energy consumption and the area overhead. In addition to the cell-level analysis, different programming schemes are also discussed in detail. Based on the study, a detailed design methodology is proposed. With the proposed methodology, designers can explore the most energy/area efficient ReRAM design with different design constraints and parameters at the very early stage of the ReRAM design.**

I. Introduction

The scaling of traditional memory technologies, such as SRAM and DRAM, is approaching its technological and physical limit. In order to effectively follow the Moore's Law [1] in the near further, new memory technologies are desired. In the past few years, the non-volatile technologies (NVM), including Phase Change RAM (PCRAM), Magnetoresistive RAM (STT-RAM) and Resistive RAM (ReRAM) have been widely accepted as candidates for 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 it be considered as the most promising technology. Researchers have shown that the state-of-art single-level-cell ReRAM can achieve sub-8ns random access time for both read and write operation with resistance ratio larger than 100 [2]. Also, HP labs and Hynix have already announced that they are going to commercialize the memristor-based ReRAM and predicted that ReRAM could eventually replace the traditional memory technologies [3].

Different from other non-volatile memory technologies, ReRAM can be implemented in a cross-point style structure without any access devices. Specifically, in a nano cross-point array, each bistable ReRAM cell is sandwiched by two layers, orthogonal nanowires, without access devices. In this case, the cell size of ReRAM can be further reduced to $4F^2$ per bit. However, the simplicity of the access device free, cross-point structure ReRAM also brings in additional challenges on the peripheral circuit design as well as the memory organization. There are many literatures that analyzed the design challenges of the cross-point ReRAM array [4]–[7]. Nevertheless, all of these researches focus on the cross-point memory array itself but do not take into account the overhead of voltage drivers and different programming methods. Besides, the analysis of area and energy consumption is also lacking. In this work, we carefully analyzed the design challenges of cross-point structure based ReRAM. A precise mathematical model is built to evaluate the reliability, energy consumption and area overhear for different design schemes and various cell parameters. Based on the study, a detailed design methodology is proposed. With the proposed methodology, designers can explore 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, the system designers can also leverage the proposed framework to provide valuable feedback to device researchers to adjust their experiments and offer more useful ReRAM cell. We believe that this kind of two-way communications will be very helpful to accelerate speed-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 built in this paper and detailed the edge conditions for different write and read schemes. Section IV analysis different design constraints of the write and read operation on the cross-point based ReRAM array. The energy consumption and area overhead are also analyzed in this section. Then in Section V, the design methodology for the ReRAM array is proposed based on our mathematical model and simulation results. Finally, the conclusion is presented in Section VI.

II. Preliminaries

This section provides preliminaries of the ReRAM technology and the cross-point architecture. Then the limitations of cross-point architecture are discussed, which motivates the work in this paper.

A. Background of ReRAM technology

Table. I ???1 compares the state-of-art non-volatile memory technologies. Obviously, the ReRAM and STT-RAM are the most promising technologies because they have faster access time than PCM and FeRAM with reasonable endurance. However, although the STT-RAM shows the fastest read/write

TABLE I
COMPARISON OF EMERGING NON-VOLATILE MEMORY TECHNOLOGIES

Metric	STT-RAM	PCM	FeRAM	ReRAM
Cell Size(F^2)	$6-20$	$4-8$	15	4
Read Latency(ns)	1-10	20-50	20-80	5-50
Write Latency(ns)	2-20	150	100	5-50
Endurance	10^{15}	10^{8}	10^{12}	10^{8-10}

latency among all non-volatile memory technologies, the structure of the memory cell is complex and it has large cell size. On the other hand, the ReRAM has very simple cell structure and can be implemented as an area efficient cross-point structure, which can work without any access devices. This simple structure provides the possibility of high density integration and 3-D stackability to ReRAM based memory array. Besides, the ReRAM has much higher ON-OFF resistance ratio than STT-RAM. Therefore, with all of these advantages, the ReRAM based memory is considered as a highly competitive technology compared to all of the emerging non-volatile memory technologies.

As implied by the name, the ReRAM uses its resistance to represent the stored information. A ReRAM cell can be switched between the high resistance state (HRS) and the low resistance state (LRS) by applying an external voltage across the cell. The resistance switching behavior of ReRAM has been noticed for several years and attracted great research interest recently for the potential application as next generation non-volatile memory technology. Generally, a ReRAM cell is usually 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 proposed by HP Labs in 2008 [8]. The proposed memristor based ReRAM is considered as the first experimental realization and a theoretical model of the fourth fundamental circuit elements, which is predicted by Chua [9] about 40 years ago. The memristor based ReRAM has a very small cell size of $50 \times 50nm^2$ with access time less than 50ns. Another H_fO_2-based bipolar ReRAM is implemented by ITRI this year with as small as 7.2ns access time [2].

Although there are different kinds of ReRAMs proposed by researchers, all of them can be classified into two types: the unipolar ReRAM and the bipolar ReRAM. For a unipolar ReRAM cell, the resistance switching behaviors do not depend on the polarity of the voltage input across the cell and only relate to magnitude and latency of the voltage input. On the other hand, for a bipolar ReRAM cell, the voltage polarity for a ON-to-OFF switching (RESET operation) is different from a OFF-to-ON switching (SET operation). A unipolar ReRAM can be easily stacked on top of a diode to built a one diode one resistor (1D1R) ReRAM [10]. However, as mentioned, since the SET and RESET operations have different latencies and therefore the performance is mainly determined by the longer voltage pulse. Besides, the control of SET, RESET and read operations without any disturbance is another crucial design challenge, especially in the high speed ReRAM design. Therefore, the reported state-of-art high performance ReRAM

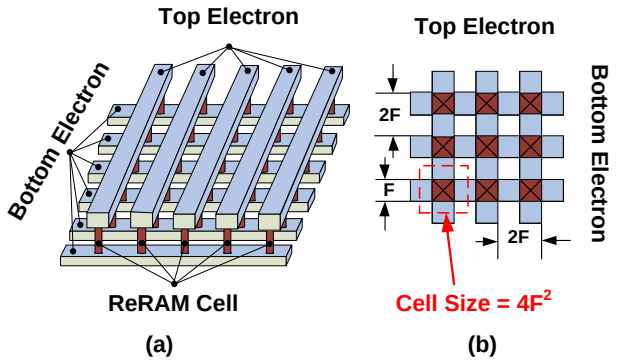

Fig. 1. A schematic view of typical cross-point architecture. (a): Over view of the cross-point architecture. (b): The layout of the cell size of $4F^2$.

technologies are dominated by bipolar ReRAM [2], [11], [12]. However, the bipolar ReRAM also has its inherent problems due to the lack of access device. Thus, in this study, we perform a detailed analysis of the design challenges of the bipolar ReRAM based cross-point array.

B. Cross-Point Architecture

There are two possible memory structures for bipolar ReRAM implementation: the traditional MOSFET-accessed structure and the cross-point structure. In the MOS-accessed memory array, a MOSFET is added as an access device for each memory cell. The size of a MOSFET access device is always much larger than size of ReRAM cell. Thus, in this case, the total area of the memory array is mainly dominated by MOSFET access device rather than the actual ReRAM cell. The ReRAM's advantage of ultra small cell size will be eliminated by the access device.

On the contrary, the cross-point structure is more area-efficient for the ReRAM based memory array [13]. A schematic view of a typical cross-point memory array is shown in Figure. 1(a). It can be seen that in the cross-point array, the only item at each crossing point is the ReRAM cell. Therefore, Figure. 1(b) shows that the cell size of the cross-point memory can achieve $4F^2$, the theoretical minimum size for a single layer single level memory cell. Besides, as aforementioned, the good stackability and the high resistance ratio provide the capability of building multi-layer multi-level cross-point ReRAM array, which can further increase the area efficiency of the ReRAM array [2] [14].

For the cross-point structure, the write operation can either write one bit per access or write several bits at the same word line at the same time. Although the second scheme is more efficient, it required a two-step writing operation to prevent the unintended writing [13]. Therefore, the write latency for the multi bit write operation is much larger than the one bit operation. Besides, the unselected word lines and bit lines can be either left floating or half biased. On the other hand, in order to read a ReRAM cell, the selected word line should be biased at a read voltage and the other word lines and bit lines are grounded. During the read operation, the current at each bit line is sensed and compared to the reference current. Then the resistance stored at the selected cell is read out. However,

due to the sneak current existed at the cross-point array, the actual current is impacted significantly by the data pattern of the unselected cells in the array. This phenomenon of read disturbance will directly restrict the size of the cross-point array. In order to mitigate the impact of sneak current, a two-step write operation can be performed: First of all, the background current of the cross-point array will be sensed. Then the total current, comprised 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. Since there are different read/write schemes as well as the array size can be chosen for the cross-point array, it is not a easy way to figure out how to design a workable memory array with the minimum energy consumption and area overhead. Thus, following sections will proposed a worst case based methodology to help designer make the choice at the early step of the design.

III. Modeling of the Cross-Point Memory

In this section, a detailed mathematical model of the cross-point array is built. By using the proposed model with specific parameters and edge conditions, the reliability, energy consumption and area overhear of different read/write schemes can be easily evaluated.

A. Basic model of Cross-Point Memory

Figure. 2 shows the circuit model of a M by N cross-point ReRAM array. The horizontal lines are word lines and vertical lines represent the bit lines. The ReRAM cells are located at each cross point of one word line and one bit line. The resistance of the ReRAM cell at the cross point of the i^{th} word line and j^{th} bit line is indicated as $R_{i,j}$. We assume the resistance of the interconnect nanowires between two adjacent cross point has the same value of R_{line}. The input resistance of each word line and bit line is R_v and the resistance of sense amplifier is R_s. In order to set up the Kirchhoff's Current Law (KCL) equations, the voltage at each cross point is indicated as $V_{i,j}$ for word line and $V'_{i,j}$ for bit line. A detailed cross point is also shown in right hand figure of Figure. 2. Besides, the input voltage for the i^{th} word line is V_{Wi} and for the i^{th} bit line is V_{Bi}. In the case of two side voltage input of word line, the voltage at the other end of the i^{th} word line is denoted as V_{W1}. Finally, the voltage at the sense amplifier is V'_{Bi} during the read operation.

B. Mathematical Model of the Cross-Point Array

Based on the circuit model shown in Figure. 2, the current equations for each cross point can be set following the KCL:

$$\Sigma_{I=1}^{k} I_k = 0. \tag{1}$$

Fortunately, all of the cross points have similar structure with no more than three current injections and therefore it is very easy to set up the KCL equation for each cross point. However, we should treat the cross points at the edges of the array seriously because there are different conditions for different

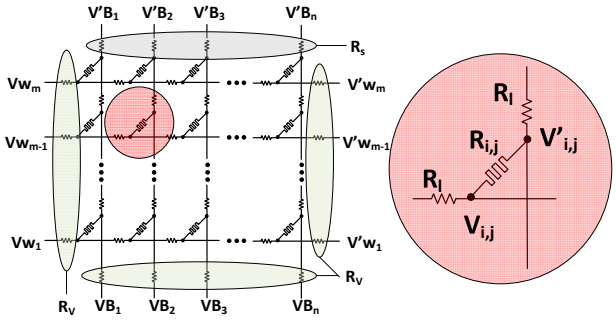

Fig. 2. The basic model of typical cross-point array.

write/read schemes. For example, the unselected word line for write operation can be either half biased or left floating. Thus, the edge conditions should be adjusted according to each write/read schemes. In particular, all of the cross points in the array can be classified into three major categories: Normal point, Activated point and Floating point.

The normal points locate 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

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

for the node at word line layer and

$$R_l^{-1}V'_{i-1,j} - (2R_l^{-1} + R_{i,j}^{-1})V'_{i,j} + R_l^{-1}V'_{i+1,j} + R_{i,j}^{-1}V_{i,j} = 0, \tag{3}$$

for the node at bit line 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. Take the point located at the intersection of i^{th} word line and 1^{st} bit line for example. If the i^{th} word line is activated by a voltage input of V_{Wi}, this cross point is an activated point, and the KCL equation for this point is:

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

Otherwise, it is floating and has the KCL equation take the form of

$$-(R_l^{-1} + R_{i,1}^{-1})V_{i,1} + R_l^{-1}V_{i,2} + R_{i,1}^{-1}V'_{i,1} = 0. \tag{5}$$

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

$$V = [{V_1}^T, {V_2}^T ... {V_m}^T, {V'_1}^T, {V'_2}^T ... {V'_m}^T]^T, \tag{6}$$

where,

$$V_i = [V_{i,1}, V_{i,2} ... V_{i,n}]^T, \quad V'_i = [V'_{i,1}, V'_{i,2} ... V'_{i,n}]^T, \tag{7}$$

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

$$A \cdot V = C. \tag{8}$$

A is a $2mn \times 2mn$ coefficient matrix, which is determined by Equation(2)-(5). C is a $2mn \times 1$ vector, containing the constant terms of these equations. As shown, the KCL equations for each node have very simple structure and are very similar to each other. Therefore, the linear equation system has a relatively fixed format and simple structure, making it very easy to establish and adjust the coefficients and constants according to different design schemes. The characteristics of the linear system can be summarized as:

1) As shown in Equation (9), the coefficient matrix A can be further partitioned into 4 smaller subblocks :

$$\mathbf{A} = \begin{bmatrix} A1 & A2 \\ A3 & A4 \end{bmatrix}. \tag{9}$$

All of these subblocks have the same size of $m \times n$. Subblock $A2$ and $A3$ are diagonal matrixes and have the value of: $A2_{i,i} = A3_{i,i} = R_{i,i}^{-1}$. Besides, $A2$ and $A3$ do not change their values with different schemas. However, $A1$ and $A4$ are a little bit more complex than $A2$ and $A3$. $A1$ is a tridiagonal matrix and only has nonzero elements at the location in the main diagonal, and the first line below and above the diagonal. Similarly, the A_4 is a special tridiagonal matrix, which has nonzero elements in the main diagonal, the n^{th} line below and above the diagonal, where n is the number of bit line in the cross point model. The value of the elements in $A1$ and $A4$ can be easily derived from Equation (2) and (3). However, as mentioned, the edge condition varies with different program schemes. Therefore, the coefficients related to the edge condition should be update according to the program schemes. Clearly, the four edges shown in Figure. 2 correspond to different coefficients in $A1$ and $A4$. Due to the space limitations, we take the nodes at the left edge of the array for example. Coefficients of other edge nodes can be initiated by the same way. The coefficients of nodes at the left edge of the array ($V_{i,1}$) can be set as:

$$A1(k,k) = \begin{cases} -(R_l^{-1} + R_{i,1}^{-1}) & \text{if } floating \\ -(R_v^{-1} + R_l^{-1} + R_{i,1}^{-1}) & \text{if } activated \end{cases} \tag{10}$$

where $k = (n-1)i + 1$ for $i = 1, 2...m$.

2) The constant terms C is a $2mn \times 1$ vector. Equation(2)-(5) show that only KCL equations of the activated points have the constant terms. Therefore, only the following elements in C may have non-zero value: $C((i-1)n+1)$, $C(in)$, $C(mn+i)$ and $C((2m-1)n+i)$ for $i = 1, 2...m$, correspond to the nodes at the four edges respectively. Likewise, we take nodes $V_{i,1}$ for example. The constant correspond to these node can be defined as:

$$C((i-1)n+1) = \begin{cases} 0 & \text{if } floating \\ -R_v^{-1} V_{Wi} & \text{if } activated \end{cases} \tag{11}$$

Therefore, with all of the required parameters, including the resistance of ReRAM cell, resistance of interconnect wires, program voltages and write/read schemes, the voltage at each cross point of the array can be obtained by solving the Equation (8) with simple matrix computations. With the detailed voltage values, $V_{2mn \times 1}$, we can analyzed the array at a very fine granularity. Also, these information can be very useful to evaluate the reliability, energy consumption, driven current density and area overhead of the cross-point array.

IV. Analysis of Design Constraints - A Case Study

In this section, a typical case of the ReRAM based cross-point array is detailed analyzed by using the proposed mathematical model.

A. Overview

As shown in Figure. 2, in order to write or read the cross-point array, the external voltages should be applied at the end of the word line and the bit line. Since there are several potential read/write schemes can be used to program the memory array, it is quite difficult to point out which scheme is the most proper choice under given design constraints of area/energy/reliability. Therefore, in this section, studies on different operation schemes and present are conducted. The requirement for array size, energy consumption and area overhead are analyzed in the worst cases scenario. The results of this study can be very useful to guide the design of the cross-point array.

Table II shows the circuit parameter of our baseline 32nm design. The data is consistent to the recently published studies on ReRAM [6] [13]. In the following the reliability, energy consumption and area overhead for the four write schemes are detailed. Then the sensitivities of these schemes to the data pattern of HRS and LRS ReRAM cells, and non-linearity are studied.

Although the purposes of read and write operation are different, both of them are realized by full biasing selected cell and floating (or half biasing) the unselected cell. Thus, the set up of the coefficient matrix A and constant vector C are very similar for the read and write operation. In addition, the energy consumption and area overhear also have similar trends for them. Therefore, in the next section, we first study the writing operation comprehensively. After that, for the read operation, we mainly focus on the read margin analysis since it is unique to the read operation.

TABLE II
Parameters of the baseline Cross-Point Array

Metric	Description	Values
S_{cell}	Cell Size	$4F^2$
R_l	Interconnection Resistance	1.25Ω
V_{RESET}	Threshold voltage for RESET	$2.0V$
V_{SET}	Threshold voltage for SET	$-2.0V$
V_{READ}	Read Voltage of Cell	$0.5V$
R_{off}	HRS Resistance	$500K\Omega$
R_{on}	LRS Resistance	$10K\Omega$
$V_W(R)$	Word Line Voltage during Read	$0.4V$
$V_W(W)$	Word Line Voltage during Write	$\pm 2V$
$V_W(H)$	Half Selected Word Line Voltage	$1V$
$V_B(R)$	Bit Line Voltage during Read	$0V$
$V_B(W)$	Bit Line Voltage during Write	$0V$
$V_B(H)$	Half Selected Bit Line Voltage	$1V$
M	Number of Word Line	64
N	Number of Bit Line	64

B. Write Operation

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

1) According the location of the selected cell, activate one word line and one bit line and leave all of the other lines floating (FWFB shemes).
2) Activate the selected word line and bit line. Left all the unselected word lines floating and half bias the unselected bit lines (FWHB shemes).
3) In contrast with the scheme 2, activate the selected word line and bit line. Left all the other bit lines floating and half bias the other wold lines (HWFB shemes).
4) Activate the selected word line and bit line. Then half bias all of the other word lines and bit lines (HWHB shemes).

Since the reliability, energy consumption and area overhead for these schemes are different from each other. We will address these problem separately and then combine all of the constraints to provide a design guideline for write operation.

Reliable Write Operation.

The most important issue for the write operation is the reliability concern. In the ideal condition, the resistances of interconnect wires and the sneak currents at unselected cells are negligible. In this case, all of these write schemes can make sure that the write voltage $V_W(W) - V_B(W)$ is fully applied across the specified cell. However, the realistic circuit is not perfect and the electronic behavior of the cross-point array will deviate from the ideal case with different data pattern stored in the ReRAM cells. A reliable write operation can be defined as: switching the selected cells into required states without disturbing the states of unselected cells. Therefore, there are two potential problems of a write operation: **write failure**, an unsuccessful write on selected cell, and **write disturbance**, an undesirable write on unselected cell. All of the write schemes should meet the reliability requirement all the time. On the other word, the designer should make sure there is not any write failure and write disturbance exist even in the worst case. Otherwise, after several unreliable write operation, the data stored at the cross-point array will become unpredictable.

First of all, we use an example to show the inherent problem of FWFB scheme, which may result in severs write disturbance. Figure. 3 shows the voltage drop across each ReRAM cell of a 64×64 cross-point array. In this example, in order to write the cell at the cross point of the 32^{th} word line and the 32^{th} bit line, the selected word line and bit line are biased at 2V and 0V, respectively. All of the other word lines and bit lines are biased at 1V. The ReRAM cells at the selected bit line are in the HRS, while all of the other cells are in the LRS. It clearly that the voltage drop across the selected cell ($V_{32,32}$) almost has the same magnitude as the unselected cell at the same bit line, resulting the write disturbance to all of the unselected cells at the selected bit line. Actually, for a $M \times N$ matrix ($M > N$), the worst case voltage drop of the

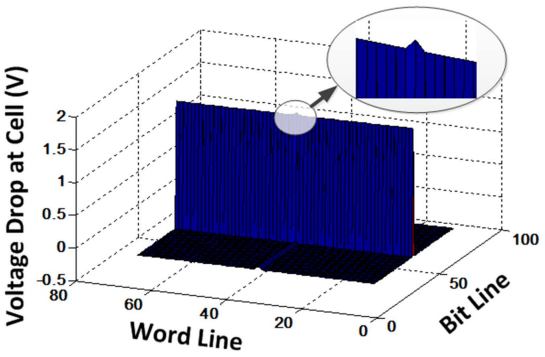

Fig. 3. Write Disturbance for FWFB Schemes. ($V_{W32} = 2V$, $V_{B32} = 0V$. $R_{x,32}$ at HRS, others at LRS.)

unselected cell can be calculated as:

$$V_{worst} = V_{select} \cdot [1 - \frac{1}{M + (N-1)R_{off}/R_{on}}]. \quad (12)$$

Considering that the reported On-OFF resistance ratio of ReRAM cell is always > 50 [2], [12], [15]–[18], , the worst case voltage drop at the unselected cell is larger than 98% of the voltage at the selected cell, making it is impossible to build a reliable cross-point structure ReRAM with the FWFB scheme. 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 the cells at one word line at the same time or only write one bit per access and separate the write operation to several arrays. In the following discussion, we start from one bit per access write operation. And then the results of one word line per access method are discussed.

The write failure mainly results from the voltage drop at the interconnect wires along the word line and bit line It has been shown that [6], for one bit per access write operation, the worst case voltage drop occurs when

$$\begin{cases} R_{M.N} = R_{on} \\ V_{WM} = V_W(W) \\ V_{BN} = V_B(W). \end{cases} \quad (13)$$

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 exceed the threshold voltage even at the worst case. Figure. 4 shows the lower bound of the driven voltage for different size of cross-point array. The minimum word/bit line voltage increases from 2.01 V for a 8×8 array to 4.47 V for a 128×128 cross-point array. Besides, for a memory capability, the cross-point array can be organized with different number of word line and bit line. For example, a 4K bits cross-point array can be implemented either by a 64×64 array or by a 32×128 array. In the latter case, the voltage drops along the word line will be much serious than along the bit line. Figure. 5 exams the voltage requirement for different array organizations with different write schemes. The result shows that from the reliability point of view, the cross-point array

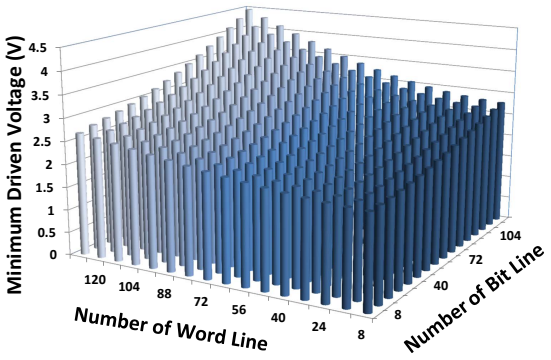

Fig. 4. Write Voltage Requirement (Threshold Voltage = 2V).

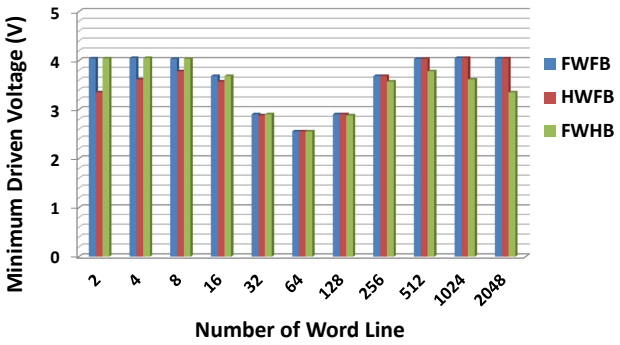

Fig. 5. Write Voltage Requirement with Different Memory Shape. (Array Capacity = 4Kbits, Activated Word Line Voltage = 2V, Activated Bit Line Voltage = 0V.)

with same numbers of word line and bit line is the best choice. Besides, we also notice that when the array has the same number of word line and bit line, FWFB, HWFB and FWHB schemes have the same minimum driven voltage.

However, boosting the driven voltage also brings in other potential problems for the array. Especially, the increasing of the driven voltage also increases the voltage applied at the unselected cell. Therefore, a **write disturbance** may occur when the voltage applied at the unselected cell exceeds the threshold voltage for SET or RESET operation. Figure. 6 shows the worst case (maximum) voltage applied at unselected cells with the minimum driven voltage shown in Figure 4. Since the threshold voltage of the ReRAM cell is 2V, only the array size with worst case voltage less than 2V are allowable. Otherwise, the array is unreliable because it can not avoid write failure and write disturbance at the same time. Therefore, Figure. 6 provides the hard constraint of 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 for a cross-point array can be calculated as:

$$E_{write} = E_{select} + E_{unselect} + E_{halfselect} + E_{line}, \quad (14)$$

where the E_{select} is the energy consumed to change the state of the selected cell, the $E_{unselect}$ and $E_{halfselect}$ are the

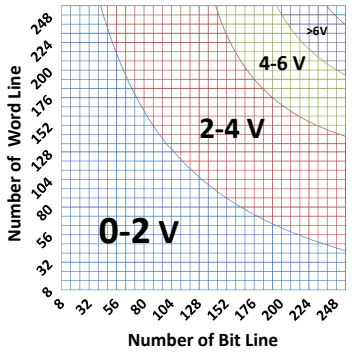

Fig. 6. The Maximum Voltage Applied at Unselected Cells with the Minimum Driven Voltage.

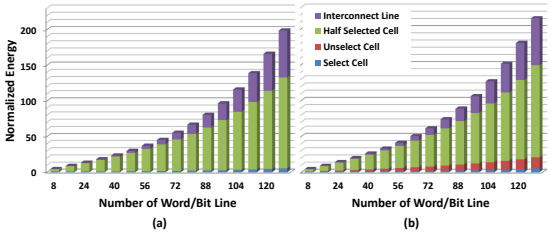

Fig. 7. The Normalized Energy Consumption. (a): HWHB scheme (b): FWHB and HWFB schemes.

undesired energy wasted at the half selected and unselected cells. The energy consumed by the interconnect lines are represented by E_{line}. Figure. 7 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. Besides, the energy wasted during the write operation takes a great part of the total energy for large array size. For example, the undesired energy consumption for writing a 128×128 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 are larger than the half-biased scheme. Due to this reason, the total energy consumptions for FWHB and HWFB schemes are at least 10% larger than that of HWHB scheme.

Area cost of Write Operation.

The write operation for a $M \times N$ array requires totaly $M + V$ voltage drivers. Therefore, the average number of ReRAM cells per voltage driver can be calculated as $mn/(m+n)$. Given the array capacity of C_{array}, it is easy to find that the optimal array organization can be achieved when $M = N = \sqrt{C_{array}}$ and the maximum number of cells per voltage driver is: $\sqrt{C_{array}}/2$. However, the area overhead of a voltage driver is also related to its current drive capability. Figure. 8 shows the maximum write current with different ReRAM array size. According to the current requirement, the area of the voltage driver can be directly calculated, which will be shown in the following discussion.

Discussion on Multi-Bits Write Operation.

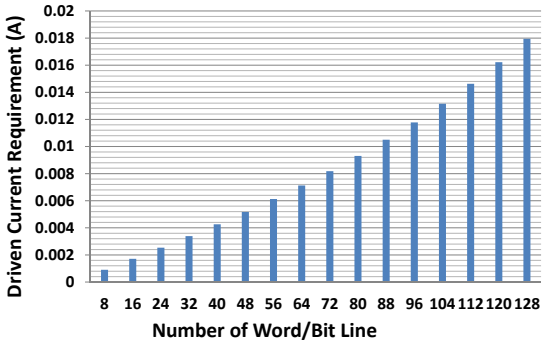

Fig. 8. The

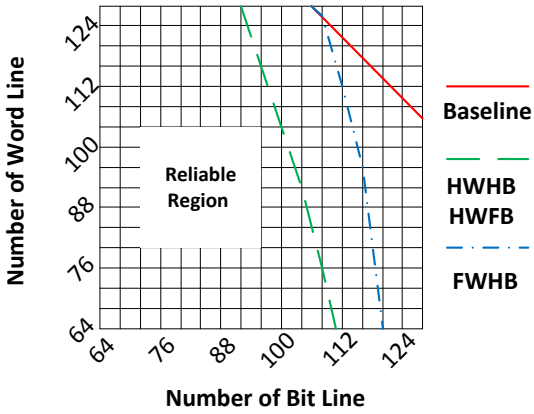

Fig. 9. The Array Size Requirement for the Cross-Point Array with Different Write Schemes. (Baseline: one bit per access. HWHB, HWFB and FWHB: one word line per access.

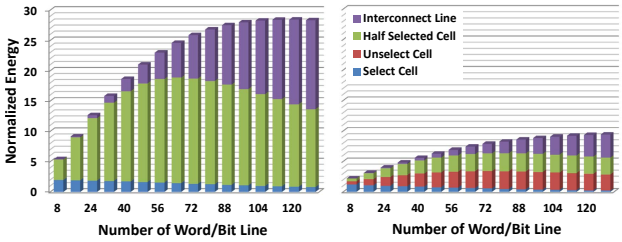

Fig. 10. The Normalized Energy Consumption per Bit for Multi-Bits Write Operation. (a): HWHB and FWHB schemes (b): HWFB scheme.

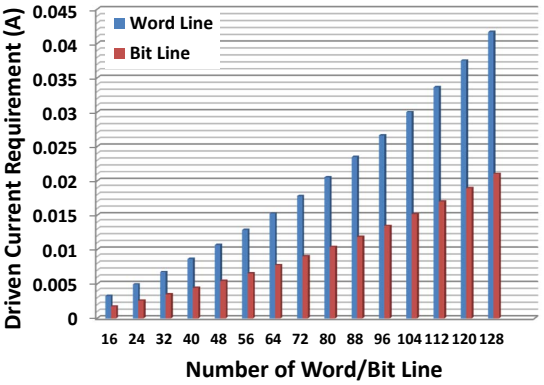

Fig. 11. The

So far, we only discuss the one bit per access write operation. At this section, the difference between one bit per access and one word line per access write operations are discussed. Firstly, writing a word line at the same time will worsen the voltage drop along the word line. Therefore, as shown in Figure. 9, 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 compared the energy consumption, we compared the energy-per-bit instead of the total energy. For example, in order to write a word line with size of 128, the energy-per-bit can be calculated as: $E_{ave} = E_{total}/128/2$. Figure. 10 shows the energy-per-bit of the multi-bit write operation. The energy shown in this figure is normalized to the same unit as Figure. 7 for easier comparison. The results show that for large size of the cross point array, the multi-bit write operation is much more energy efficiency. This is because the energy wasted at the unselected and half-selected cells are shared by multi bits and the average energy for one each bit is therefore reduced. However, although the multi bit write operation has the advantage on energy consumption, the maximum current requirement for each word line also increases. As shown in Figure. 11, the maximum driven current for bit line is almost the same as one bit writing, the drive capability for wold line is almost doubled for multi-bit writing. Since the area of the voltage driver increases proportional with its driven capability, the area overhead for multi-bit writing is about 50% larger than one bit writing.

Non-linearity of the ReRAM Cell.

One of the most distinct feature of ReRAM is it non-linearity. Take the memristor based ReRAM for example, the non-linearity can be described as: 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 equivalent resistance of the memristor biased at V/p and V [13]. Normally, the $K_r(p,V)$ value for memristor based ReRAM is larger than 5, meaning that the resistance of half-biased cell is 10 times larger than full-biased cell. Clearly, the ReRAM cell with larger non-linearity coefficient is more suitable to be employed as the memory cell since the current in the sneak path will be significantly reduced. Besides, the increased resistance at half-selected and unselected cell can also mitigate the voltage drop along the activated word line and bit line. Figure 12 shows the influence of different non-linearity coefficients on the array size requirements. Due to the space limitation, only the results for one bit writing with HWHB scheme are shown. In this figure, the maximum array size increases from 112×112 to 340×340 when the non-linearity coefficient K_r increases to 10.

On the other hand, the non-linearity can also benefit the energy consumption and area overhead as well. Take the $128\times$

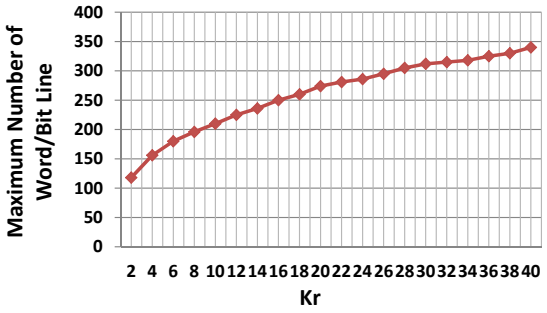

Fig. 12. The Maximum Array Size with Different Non-linearity

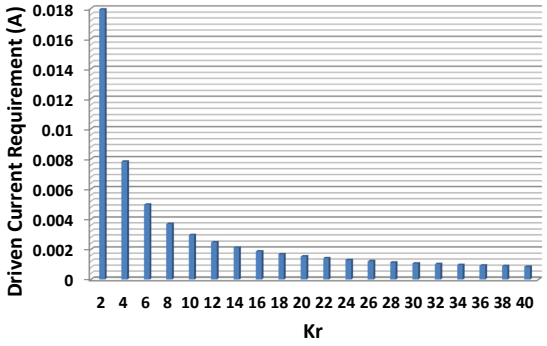

Fig. 13. The

128 array for example.

C. Read Operation

The design constraints for read operation can be analyzed by the same way as the write operation, including the reliability, energy consumption and the area overhead. Note that the read voltage/current is much lower than the write voltage, therefore we believe the voltage drivers can always provide enough current for the read operation if they meet the current requirement for write operation. Thus in this section, we only shows the results of reliability issue and read energy. Also, since the difference

The reliability issue of the read operation of the cross bar array are well studied in previous researches.

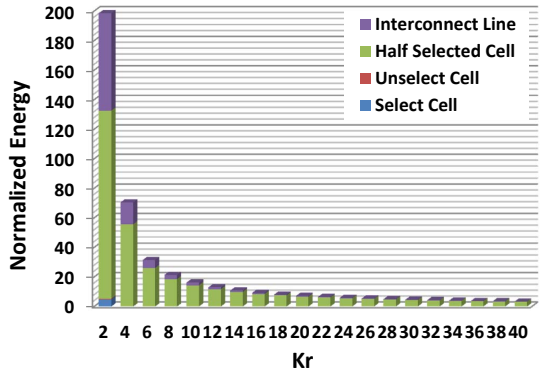

Fig. 14. The

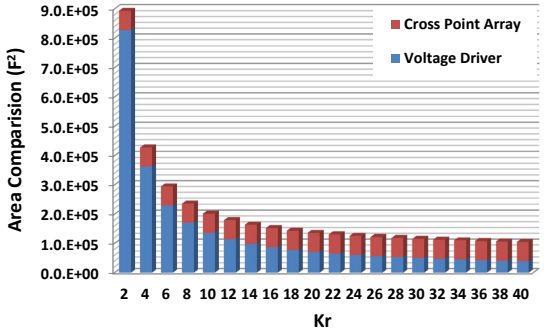

Fig. 15. The

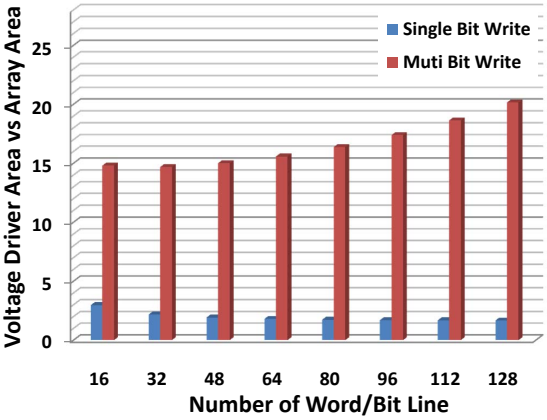

Fig. 16. The

V. Design Methodology

Based on the analysis of Section IV, a ?????

At the initialization step, the physical parameters and the design constraints are firstly input. Based on physical parameters the coefficients matrix A_{basic} and the vector of constant term C_{basic} are initialized. Since the A_{basic} and C_{basic} do not change with the different

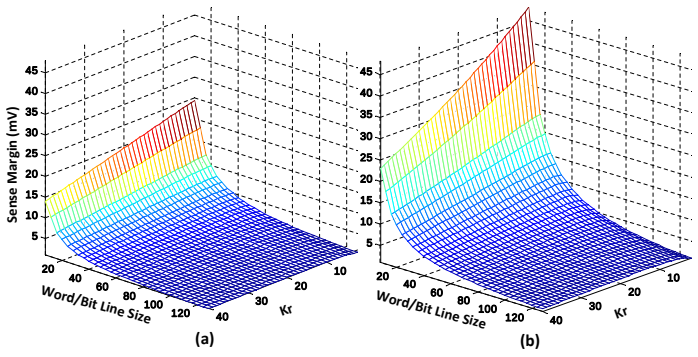

Fig. 17. The

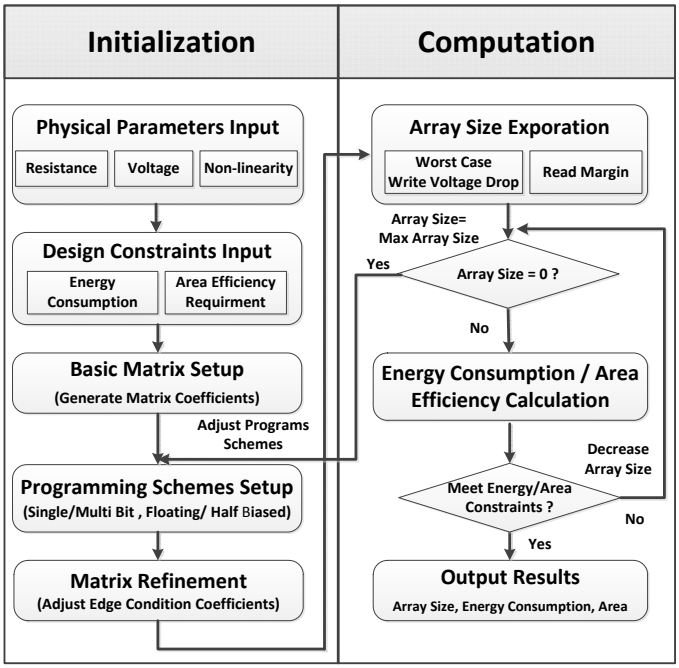

Fig. 18. The

VI. Conclusion

The ReRAM is a promising candidate of the next-generation non-volatile memory technology. The area efficient cross-point structure is the most attractive memory organization for the ReRAM based memory design. However, intrinsic problems of the cross-point structure, such as the existence of sneak current and the voltage drop along the nanowire bring in extra challenges to the design of reliable ReRAM based memory array. In this paper, a mathematical model for the cross-point array is proposed. We show that the propped model has a vary simple structure and is flexible to evaluate different write/read schemes. By using this model, the design constraints, including the array size, energy consumption and area overhead, are analyzed in details. Based on the results of our study, a detailed design methodology is proposed, which can help designers explore the most energy/area efficient ReRAM design with different design constraints and parameters at the very early stage of the ReRAM design.

References

[1] G. E. Moore, "Cramming more components onto integrated circuits," *Electronics*, vol. 38, no. 8, April 1965.

[2] 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.

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

[4] 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.

[5] 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.

[6] 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.

[7] 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.

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

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

[10] M. Lee *et al.*, "2-stack 1d-1r cross-point structure with oxide diodes as switch elements for high density resistance ram applications," in *Electron Devices Meeting, 2007. IEDM 2007. IEEE International*, Dec 2007, pp. 771 –774.

[11] 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.

[12] 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.

[13] 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.

[14] 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.

[15] C. Ho *et al.*, "9nm half-pitch functional resistive memory cell with $<$1ua programming current using thermally oxidized sub-stoichiometric wox film," in *Electron Devices Meeting (IEDM), 2010 IEEE International*, Dec 2010, pp. 19.1.1 –19.1.4.

[16] W. Chien *et al.*, "A forming-free wox resistive memory using a novel self-aligned field enhancement feature with excellent reliability and scalability," in *Electron Devices Meeting (IEDM), 2010 IEEE International*, Dec 2010, pp. 19.2.1 –19.2.4.

[17] J. Lee *et al.*, "Diode-less nano-scale zrox/hfox rram device with excellent switching uniformity and reliability for high-density cross-point memory applications," in *Electron Devices Meeting (IEDM), 2010 IEEE International*, Dec 2010, pp. 19.5.1 –19.5.4.

[18] H. Lee *et al.*, "Evidence and solution of over-reset problem for hfox based resistive memory with sub-ns switching speed and high endurance," in *Electron Devices Meeting (IEDM), 2010 IEEE International*, Dec 2010, pp. 19.7.1 –19.7.4.