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 Memory (ReRAM) have attached much attention because their non-volatility, high speed, low power consumption and good scalability. Among these emerging memory technologies, the ReRAM has shown great potentials a one of the most promising candidates for future universal memory is the ReRAM, 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 with out 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 comprehensively analyzed. In addition to the cell-level analysis, ?????????????. A precise mathematical model is built to perform ..... Based on the study, a detailed design methodology is developed. With the proposed methodology, designers can explore the most energy/area efficient ReRAM design with different design constraints.**

I. Introduction

The scaling of traditional memory technologies, such as SRAM and DRAM, is approaching a technological and physical limit. In order to effectively follow the Moore's Law [?] in the near further, new memory technologies are desired. In the past few years, the non-volatile technologies, including Phase Change RAM (PCRAM), Magnetoresistive RAM (STT-RAM) and Resistive Memory (ReRAM) have been widely accepted as the candidates for next generation memory 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 has 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 > 100 [?]. 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 [?].

Different from other non-volatile memory technologies, ReRAM can be implemented in a cross-point style structure. Generally speaking, in a nano cross-point array, the bistable ReRAM cell is sandwiched by two layers, orthogonal nanowires, without any access device. Therefore, the cell size of ReRAM can be further reduced to $4F^2$ per bit. However, the simplicity of cross-point structure with out access cell also brings in additional challenges on the peripheral circuit design as well as memory organization. There are many literatures that analyzed the design challenges of the cross-point ReRAM array [?] [?] [?] [?]. Nevertheless, all of these researches focus on the cross-point memory array itself but do not take into account the peripheral circuitry 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 of different design schemes and various cell parameters. Based on the study, a detailed design framework is developed. 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 **??**, the memristor based memory and the basic concept of ECC are introduced. Section **??** discusses the mathematical model used in this paper and evaluates various ECC designs for memristor-based ReRAM architecture. Section **??** shows results of the experiment conducted in this study. Finally, the conclusion is presented in Section **??**.

II. Preliminaries and Motivations

This section provides some preliminaries on ReRAM technology and 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. **??** 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 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 a cross point structure, which can work without access devices. The easy structure provides the possibility of high densigy integration and 3-D stackability to ReRAM based memory. Besides, the ReRAM also have much higher ON-OFF resistance ratio than STT-RAM. Therefore, with all of this advantages, ReRAM based memory is a highly competitive technology compared to all of the emerging non-volatile memory technologies.

As implied by the name, the ReRAM use its resistance to represent the stored information. The resistance of a ReRAM cell can be

TABLE I
Comparison of 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}

switched between high resistance state (HRS) and low resistance state (LRS) by applying an external voltage across the cell. The resistance switching behavior of the metal oxides have been noticed for several years and attracted great research interest recently for the potential application as next generation non-volatile memory technology. A ReRAM memory cell is usually built on a Metal-Insulator-Metal (MIM) structure. The resistance switching behaviors have been observed in many MIM nanodevice with different metal oxide materials. For example, a TiO_2 based MIM structure ReRAM was proposed by HP Labs in 2008 [**?**]. The proposed ReRAM is considered as the first experimental realization and a theoretical model of the fourth fundamental circuit elements, which is predicted by Chua [**?**] 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 HfO_2-based bipolar ReRAM is implemented by ITRI this year with as small as 7.2ns access time [?].

Although there are several different ReRAM proposed by researchers, all of them can be divided into tow classes: 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. However, for a bipolar ReRAM, 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 diodes to built a one diode one resistor (1D1R) ReRAM. However, as mentioned, the SET and the RESET operation have different latency and therefore the performance is mainly determined by the long voltage pulse. Besides, the control of SET, RESET and read operation without any disturbance is another crucial design challenge, especially in the high speed ReRAM design. Therefore, the reported state-of-art high performance ReRAM technologies are dominated by bipolar ReRAM [? Added reference here].

B. Cross-Point Architecture

There are two possible memory structures for ReRAM implementation: the traditional MOSFET-accessed structure and the cross-point structure. For the one cell one access device structure. In the MOS-accessed memory array, the conventional memory cell is substituted by the ReRAM cell where the access device remains to be the MOSFET. In this structure, each ReRAM cell has to be accompanied with a MOSFET access device, whose size is much larger than the ReRAM cell. In this case, the area of the memory array is mainly dominated by MOSFET access device rather than the actual ReRAM cell. Therefore, the ReRAM's advantage of ultra small cell size will be eliminated.

On the contrary, the cross-point structure is more area-efficient for the ReRAM based memory array [**?**]. A schematic view of 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, the area of the array is significantly reduced since the large MOSFET access part is removed. 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 [?] [**?**].

For the cross-point structure, a two-step writing methodology, ERASE-before-RESET, is used to prevent the unintended writing. In read operation two ways are exhibited for preventing a read failure: the first is to supply the same voltage to the unselected row and selected column. In this way, only the data on the select row is read from the selected column. The disadvantage of this method is the voltage drop on the crossing points of the unselected row and the selected column may not be ideal zero because of variations, and this imposes a limitation on the array size. The second way is a two-step write operation. The disturbance current of the partial selected cell on the selected column will be read out beforehand as a background current. Later the total current, comprised of both partial selected cell and full 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.

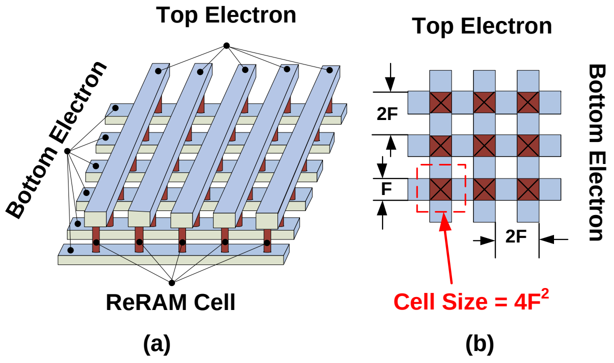

Fig. 1. A schematic view of typical cross point architecture.

III. Modeling of the Cross-Point Memory

In this section, a detailed mathematical model of cross-point memory array is built. By using the proposed model with specific parameters and boundary conditions, different read/write schemes can be easily evaluated at the very early stage of the design.

A. Basic model of Cross-Point Memory

Figure. 2 shows the circuit model of the M by N cross point ReRAM array. The horizontal lines are word line and vertical lines represent the bit line. 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 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. Edge Conditions for Different Write/Read Schemes

Based on the circuit model, the current equation for each cross point can be built following the KCL:

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

All of the cross points have similar structure and therefore it is easy to set up the KCL equation for each cross point. However, the cross point at the edge of the array may have different condition for different write/read schemes. For example, the unselected word line

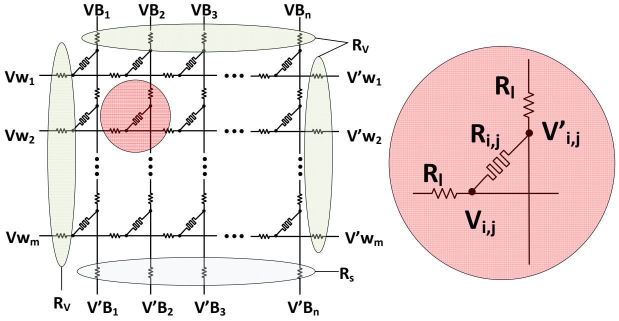

Fig. 2. The basic model of typical cross point array.

for write operation can be either half biased or left floating. Thus, the edge conditions should be carefully considered for each write/read schemes. However, generally speaking, all of the cross points can be classified into three major categories: Normal point, Activated point and Floating point.

The normal point located insides 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

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

for the node at word line layer and

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

for the node at bit line layer.

Besides, the activated point and floating point represent the node at the edge of cross point array with different conditions: a edge point, which have been directly connected to the voltage input or the ground, can be considered as a activated mode. Otherwise, it is floating node. 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}, then this cross point is activated point, and the KCL equation for this point is:

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

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

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

For the reasons of clarity, a vector $V_{2mn\times 1}$ is used to represent all of the variables in the KCL equations:

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

where,

$$V_i = [V_{i,1}, V_{i,2} ... V_{i,n}]^T, \quad (7)$$

$$V'_i = [V'_{i,1}, V'_{i,2} ... V'_{i,n}]^T. \quad (8)$$

Then the KCL equations can be considered as a system of linear equations and has the form of

$$A_{2mn\times 2mn} \cdot V_{2mn\times 1} = C_{2mn\times 1}, \quad (9)$$

where A is the coefficient matrix of and C contained the constant terms. As shown in Equation(**??**)-(**??**), the KCL equations for each node have very simple structure and are very similar to each other. Therefore, the system of linear equations has a relatively fixed format and simple structure, which will be very easy to establish the questions and to adjust the coefficient according to different design schemes. As shown in Equation (refequ:blockedmatrix), the coefficient matrix A can be partitioned into 4 smaller blocks.

$$\mathbf{A} = \begin{bmatrix} A1 & A2 \\ A3 & A4 \end{bmatrix} \quad (10)$$

All of the submatrix has the same size of $m \times n$. $A2$ and $A3$ are diagonal matrixes and have the form of: $A2_{i,j} = A3_{i,j} = R_{i,j}^{-1}$. $A1$ and $A4$ are a little bit more complex than $A2$ and $A3$. $A1$ can be further divided into

is a tridiagonal matrix and only has nonzero elements at the location in the main diagonal, and the first line below and above diagonal. Similarly, the A_4 is a special tridiagonal matrix, which has nonzero elements in the main diagonal, and the n^th line below and above diagonal, where n is the number of bit line in the cross point model. The value of the elements in A_1 and A_4 can be easily derived from Equation (**??**) and (**??**). 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 scheme as follow:

detailed discuss on the mathematic

C. Analysis of the Computing Complexity

IV. Analysis of Design Constraints

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. Although there are several potential read/write schemes can be used to program the memory array, it is difficult to point out which schemes 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 results of this study can be very useful to guide the design of the cross point array. Since it is impossible to consider all of the data pattern stored in the array, in this work, the best and worst cases are studied.

Table **??** shows the circuit parameter of our baseline design. The data is consistent to the recently published studies on ReRAM [?] [?].

Considering that program schemes for write and read operation are different and the the requirement for write and read are also dissimilar, in the following section we carefully study the write and read operation separately. And then the results are combined together to provide a design methodology for the cross point array.

TABLE II
Parameters of the baseline Cross Point Array

Metric	Description	Values
S_{cell}	Cell Size	$4F^2$
R_l	Interconnection Resistance	1.25Ω
R_s	Resistance of SA	100Ω
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	$\pm 1V???$
$V_W(W)$	Word Line Voltage during Write	$0/2V$
$V_W(H)$	Half Selected Word Line Voltage	$1V$
$V_B(R)$	Bit Line Voltage during Read	$10K\Omega????$
$V_B(W)$	Bit Line Voltage during Write	$0/2V$
$V_B(H)$	Half Selected Bit Line Voltage	$1V$
M	Number of Word Line	64
N	Number of Bit Line	64

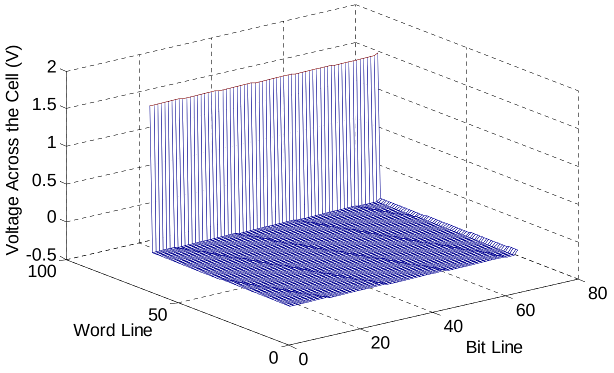

Fig. 3. Write Disturbance for FWFB Schemes.

B. Write Operation

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

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

In the ideal condition, all of these four schemes can provide enough voltage drop across the specified cell. However, the realistic circuit is not perfect and the electronic behavior of the array will deviate from the ideal scenario. In this section, the reliability, energy consumption and area overhead for the four write schemes based on our basic circuit parameter as shown in Table. **??**. Then the sensitivities of these schemes to resistance values of HRS and LRS ReRAM cell, and interconnect wire are studied.

Reliable Write Operation.

The most important issue for the write operation is the reliability concern. A reliable write operation can be defined as: switching the selected cells into required states without disturbing the states of unselected cells. Therefore, there exist two potential write error: **write failure** and **write disturbance**. All of the write schemes should at least meet the reliability requirement at the worst case. On the other word, the designer should make sure there is not any write failure and write failure disturbance occur even in the worst case. First of all, we will show the inherent problem of FWFB scheme, which may result in severs write disturbance. The worse case scenario for FWFB write disturbance can be defined as: all of unselected cells in the activated word line (or all of unselected cells in the activated bit line) are at HRS and other cells are in LRS. In this case, as shown in Figure.

The worst case scenario of write disturbance for each each schemes can be

1. Only consider the one-side scheme Consider 1/2 1/3 and floating? Energy Issue? Define Energy Efficient Parameter? Reliable Issue: Any possibility of write disturbance

2. Present the voltage drop problem Reliability Issue
3. Two Way Scheme

C. Read Operation

D. Framework

E. Experimental Results

F. Conclusion