A Novel Peripheral Circuit for RRAM-based LUT

Yi-Chung Chen and Hai (Helen) Li
Department of Electrical and Computer Engineering
Polytechnic Institute of NYU, Brooklyn, NY, USA
Email: ychen33@students.poly.edu, hli@poly.edu

Wei Zhang
School of Computer Engineering
Nanyang Technological University, Singapore
Email: zhangwei@ntu.edu.sg

***Abstract*— Resistive random access memory (RRAM) is a promising candidate to substitute static random access memory (SRAM) in lookup table (LUT) design for its high density and non-volatility. RRAM cells are fabricated at backend CMOS process and have negligible area cost. However, the complex peripheral circuit design to satisfy performance and accuracy requirements becomes a major issue. In this work, we propose a novel peripheral circuit for RRAM-based LUT. A new decoding scheme that supports dynamic programming is introduced. Furthermore, the expanded RRAM crossbar array together with the latch comparator based sense amplifier can dramatically reduce design complexity, decrease area cost, and improve tolerance on process variations. Compared to a 6-input SRAM-based LUT, the proposed RRAM-based one cuts off 60.4% of layout area. The maximal operating frequency reaches 1GHz at 10mV input difference. Simulations also show that the proposed LUT functions properly even RRAM resistances deviates 20% from the design value.**

I. Introduction

Lookup table (LUT) is the basic reconfigurable logic element used in *field programmable gate array* (FPGA) [1] and *complex programmable logic device* (CPLD) [2]. The LUTs in commercial products such as Stratix [1] and Vertex [3] are built with *static random access memory* (SRAM) cells. Since data in SRAM cannot be retained without power supply, such a design uses Flash memory to store the logic configuration and requires initialization (*i.e.*, programming LUTs) at the beginning of each powering up period. ROM-based and antifuse-based LUT designs do not suffer from data loss during powering off, however, they can be programmed only once.

Utilizing the emerging nonvolatile memory technologies in FPGA design can reduce the area of LUT, remove the extra Flash memory, and lower die size and fabrication cost. Previously, Paul and Bhunia proposed a *spin transfer torque RAM* (STT-RAM) based hybrid architecture [4]. However, it still pays high area cost on memory cells that use NMOS transistor as switching device. Nanowire-based nanoFPGA/nanoPLA can further enhance the integration density and area efficiency [5] though its fabrication process needs to be improved.

Resistive RAM (RRAM) could be another promising candidate in reconfigurable system because of its nonvolatility, high density, low power consumption, and good scalability [6]. An RRAM-based LUT consumes a small portion of area on memory cells. But reducing the design complexity and area of peripheral circuitry while maintaining the required accuracy and access latency, especially in read operations, is challenging. To solve this critical issue, we propose a novel peripheral circuit design for RRAM-based LUT. Our main contributions are summarized as follows:

- We adopted a compact decoder design, rather than the conventional multiplexer in SRAM-based LUT, to enable bit-addressability and increase configuration flexibility.
- We proposed an expanded RRAM crossbar array with an extra column of reference cells that can mitigate sensing margin degradation induced by sneak paths and remove external reference generation.
- We designed a latch comparator based sense amplifier which can work at up to $1\mathrm{GHz}$ frequency when input difference is more than $10\mathrm{mV}$.

To demonstrate the effectiveness of the proposed design, we set the high and low ratio of RRAM device $\mathrm{R_H}/\mathrm{R_L} = 5$, which is much smaller than the normal device specification. Our simulation result shows that a 6-input RRAM-based LUT can save 60.4% of area compared to the conventional SRAM-based LUT. At $500\mathrm{MHz}$ frequency, the read access latency is $< 400\mathrm{ps}$. The proposed LUT can tolerate variation up to 20% of device resistances. As $\mathrm{R_H}/\mathrm{R_L}$ increases, the enhancements on area and variance tolerance become even higher.

II. Background

LUT is a simple and popular logic element used in modern FPGAs, which is usually constructed by SRAM memory to store logic configuration. As shown in Fig. 1(a), the inputs of a LUT determines the target memory cell, whose access is controlled by *multiplexer* (*e.g.*, 16:1 MUX) [7]. The LUTs in FPGAs are sequentially programmed every time when the chip is powered up. This procedure is called as *initialization*. The conventional LUT does not support local addressing for dynamic programming. The whole chip shares one SRAM programming circuit for initialization. Some modern FPGAs provide dynamic configurable LUTs by embedding address decoding and memory configuration circuits inside LUTs [8]. However, the extra circuits induce significant area overhead therefore only a portion of a FPGA chip can adopt the dynamic configurable LUTs.

In general, RRAM denotes all the random access memories that rely on the resistance differences to store data. Various switching mechanism of RRAM has been studied [9]. As we shall show in Section III-B, our design can be applied to those bipolar RRAM devices whose resistances are inversely proportional to device dimension. Many RRAM materials, *e.g.*, ECM and PMC [10][11][12], demonstrate this property. To be

978-1-4673-0219-7/12/$31.00 ©2012 IEEE

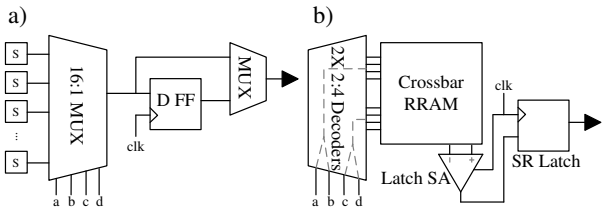

Fig. 1. (a) SRAM-based LUT. (b) The proposed RRAM-based LUT.

more conservative, we set the resistance of *low resistance state* (LRS) as $R_L = 20K\Omega$ and that of *high resistance state* (HRS) as $R_H = 100K\Omega$ in this work.

III. The Proposed Read Scheme

Fig. 1(b) depicts the architecture of a corresponding RRAM-based LUT. The crossbar array structure requires two-dimensional control on both horizontal and vertical directions. Moreover, a sense amplifier is needed for small signal detection. A detail diagram of the proposed read scheme for RRAM-based LUTs is shown in Fig. 2. During a read operation, only one cell (*e.g.*, the blue cell) is accessed by raising up the voltage on its wordline W_1 and turning on the corresponding selector of its bitline B_4. Those unselected wordlines are forced to ground while the unselected bitlines are floating. The design details are described in this section.

A. RRAM Crossbar Access Control

The conventional SRAM-based LUT can simply use multiplexer to control memory access [7]. Such a simple scheme is not optimal for crossbar array structure, which requires two-dimensional control on both wordline and bitline. Fig. 2 shows that the proposed array access control scheme includes decoders, wordline drivers, and bitline selectors.

The yellow block in Fig. 2 is the wordline driver circuitry, which supplies a low V_{read} to the selected wordline and grounds all the unselected wordlines. The transistors MP_{DR} and MN_{DR} are shared by all the wordline drivers. The read enable signal (RE) is used to turn on the driving circuitry or to force it into power save mode.

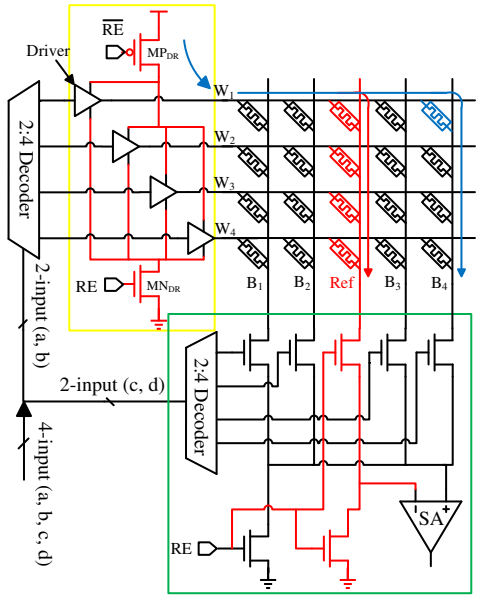

Fig. 2. The proposed read scheme.

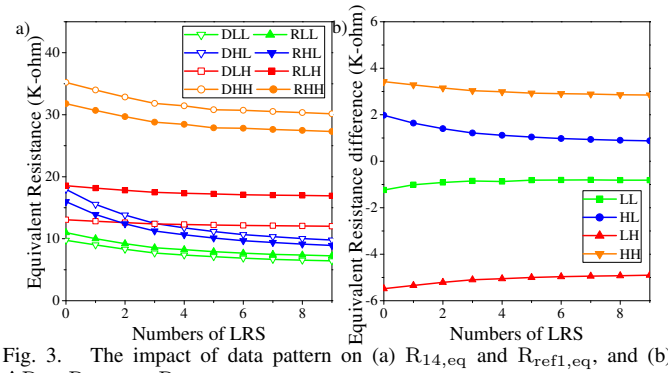

Fig. 3. The impact of data pattern on (a) $R_{14,eq}$ and $R_{ref1,eq}$, and (b) $\Delta R = R_{14,eq} - R_{ref1,eq}$.

The bitline selector circuitry is shown in green block of Fig. 2. A LUT exports only one bit data each time. Therefore, we can use NMOS transistors to select one bitline and block the others. The sensing current flows through crossbar, the NMOS selecting transistor, to the NMOS load. Since only one bitline is activated, the sensing current concentrates along the driving path (highlighted in blue) and conducts a larger voltage across the NMOS load.

This design integrates the address decoders inside LUT so as to support array bit-access and dynamic reconfiguration.

B. Expanded RRAM Crossbar Structure

1) Impact of Data Pattern on Crossbar Array: The data pattern of a crossbar array has a great impact on the sensing margin due to the existence of sneak path [13]. For example, when reading data from the target cell R_{14} in Fig. 2, the corresponding equivalent resistance between wordline W_1 and bitline B_4 ($R_{14,equ}$) is determined by not only R_{14} itself but also the resistances of all the other RRAM devices in the array.

The hollowed symbols in Fig. 3(a), whose labels start with "D", represent $R_{14,eq}$ under different data patterns. The second letter in the label is the resistance state of R_{14} – LRS (L) or HRS (H). The third letter represents that the other cells along the driving path (highlighted in blue) are all in LRS or HRS. The $x-axis$ is the number of memory cells in LRS in the crossbar except target data cell and cells along driving path. Under the different data patterns, $R_{14,eq}$ varies in a large range. Therefore, it is difficult and even impossible to find an external reference to distinguish the HRS and LRS of the target cell.

2) Expanded Crossbar Array: Some previous crossbar array designs use a specific current or voltage as reference to detect the stored data [14]. These extra circuitries require complex control and result in high area overhead, and thereby, cannot be adopted in LUT. Instead, we proposed an expanded crossbar array by introducing an extra column of reference cells, as shown in Fig. 2.

In our design, the resistance value of the reference cells are set as $(R_H + R_L)/2$. We observed the inversely proportional relationship between device resistance and footprint in many RRAM materials [10][11][12]. So the intermediate resistance value of a reference cell can be achieved by enlarging the device footprint and programming it to HRS.

The solid symbols in Fig. 3(a), labelled as "R**", represent $R_{ref1,eq}$ under the data patterns as R_{14}, denoted as "D**".

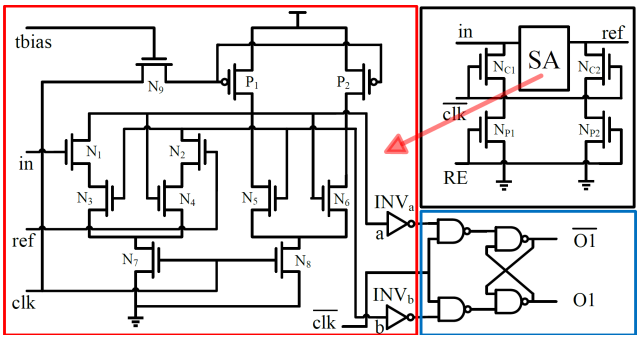

Fig. 4. The proposed sense amplifier in RRAM-based LUT.

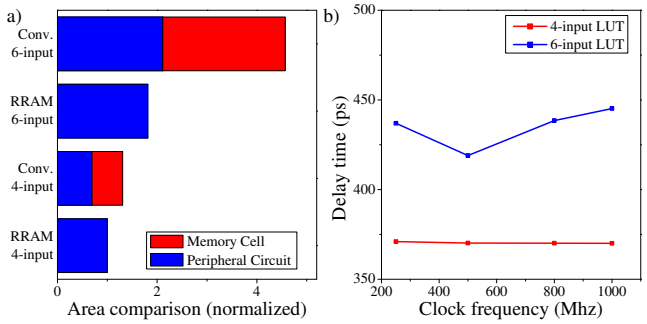

Fig. 5. (a) LUT area comparison. (b) Access latency of the proposed RRAM-based LUT at various operating frequencies.

The simulation results show that $\mathrm{R_{ref1,eq}}$ follows the similar trend as $\mathrm{R_{14,eq}}$, that is, the data pattern has the similar impact on both the target cell and its corresponding reference cell. Furthermore, we show $\Delta \mathrm{R} = \mathrm{R_{14,eq}} - \mathrm{R_{ref1,eq}}$ under different data patterns in Fig. 3(b). When the target cell is at HRS (LRS), ΔR is always positive (negative) and maintains enough difference for sensing scheme.

C. Sense Amplifier

RRAM crossbar array usually is equipped with complex sense amplifiers to maintain a high sense accuracy and small access latency. Such designs cannot be adopted in LUT. In this work, we proposed a novel sense amplifier scheme by bringing in latch comparator [15] as shown in Fig. 4.

1) Latch Comparator: The latch comparator for input signal comparison is the most important component of the proposed sense amplifier design. Its detail schematic is shown in the red block in Fig. 4. NMOS transistors $\mathrm{N_7}$ and $\mathrm{N_8}$ are used to turn on/off the latch comparator: when clock signal (CLK) is '1', the circuit is in sensing phase; otherwise, the latch comparator enters non-sensing phase.

In the sensing phase, $\mathrm{N_1}$ and $\mathrm{N_2}$ capture the input signals, *i.e.*, in and ref generated by data and reference cells, respectively. Decreasing the widths of $\mathrm{N_1}$ and $\mathrm{N_2}$ can result in smaller parasitic capacitance and fast operating speed. $\mathrm{N_3}$, $\mathrm{N_4}$, $\mathrm{N_5}$, and $\mathrm{N_6}$ together form a feedback loop. PMOS transistors $\mathrm{P_1}$ and $\mathrm{P_2}$ are active loads to amplify the latched signals. External signal $\mathrm{V_{tbias}}$ is used to control the resistance of $\mathrm{N_9}$ and hence tune the active loads – $\mathrm{P_1}$ and $\mathrm{P_2}$. We can optimize sensing latency by controlling $\mathrm{V_{tbias}}$. Although the latch comparator can generate rail-to-rail output signals, we added two inverters $\mathrm{INV_A}$ and $\mathrm{INV_B}$ at the complementary outputs in order to reduce sensing latency, improve rising/falling slope, and increase driving ability.

2) Noise Cancellation Circuit: Noise interference has a significant impact on the latch comparator's operation. We built the noise cancellation circuit surrounding the latch comparator. $\mathrm{N_{C1}}$ and $\mathrm{N_{C2}}$ are added to reduce the impact of kick-back noise from the latch comparator. In non-sensing phase, both in and ref are tied to ground. Moreover, power save mode can be activated by $\mathrm{N_{P1}}$ and $\mathrm{N_{P2}}$ controlled by RE.

3) SR Latch: Modern LUT design usually use D flip-flop for output synchronization. Considering that the proposed sense amplifier includes a latch comparator, we simply utilize an SR latch at the outputs of the latch comparator to realize the function of D flip-flop. The SR latch is controlled by the inverted clock $\overline{\mathrm{clk}}$. This design can balance the capacitive loads at the outputs of the latch comparator and reduce layout area.

IV. Simulation Results

We compare the proposed RRAM-based LUT with the conventional SRAM-based design, and evaluate the performance of its peripheral circuits. The peripheral circuits were developed by using TSMC $0.18\mu\mathrm{m}$ technology.

A. Area and Delay

Area cost is one of the major concerns in LUT design. Fig. 5(a) compares the layout areas of the proposed RRAM-based LUT and the conventional SRAM-based LUT. Here, 4×5 and 8×9 RRAM crossbar arrays are used to construct 4-input and 6-input LUTs, respectively.

In the conventional design, SRAM memory occupies more than 50% of overall area. On the contrary, the peripheral circuitry dominates the area in RRAM-based LUT since RRAM crossbar can be built on top of CMOS process. A 4-input RRAM-based LUT, which includes decoders for dynamic configuration, achieves 23.6% area saving compared to the SRAM-based one without supporting dynamic configuration. The extra decoding circuit consumes $\sim$ 23% of area in RRAM-based design. When LUT's input number increases from 4 to 6, the area of a RRAM-based LUT is only 39.6% of that of a SRAM-based one.

The LUT design can function well at up to 1GHz clock frequency. Fig. 5(b) shows the access latency of the proposed read scheme when varying operating frequency. The given sense amplifier design was fine tuned at 500MHz. So the best performance at this frequency is observed. The access latency of the sense amplifier at the other operating frequencies can be improved by carefully adjusting $\mathrm{V_{tbias}}$. When increasing LUT input number from 4 to 6, the access latency increases from $\sim$ 370ps to $\sim$ 435ps due to the increased capacitive load induced by bigger array. Comparably, the access latency of a conventional SRAM-based LUT at TSMC $0.18\mu\mathrm{m}$ technology is in the range from $\sim$ 500ps to $\sim$ 700ps.

B. Sense Amplifier

Fig. 6(a) shows the waveforms of the proposed sense amplifier at 500MHz frequency. The nodes are corresponding to those in the schematic of Fig. 4. To demonstrated the effectiveness of design, we utilized the critical data pattern

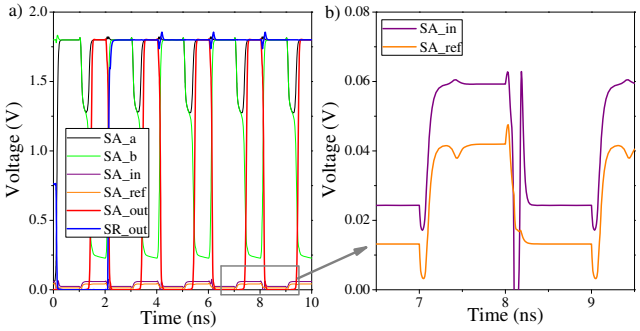

Fig. 6. (a) The simulated waveforms of sense amplifier in 4-input LUT under critical sensing condition of data pattern. (b) The enlarged view of sense amplifier inputs.

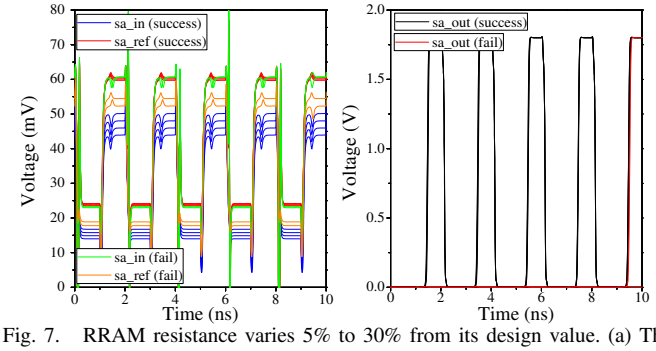

Fig. 7. RRAM resistance varies 5% to 30% from its design value. (a) The waveforms of sense amplifier inputs under the worst case data pattern. (b) The corresponding sensing result.

when the target cell and all the other cells are in LRS resulting in the smallest sensing margin. The signal difference at the inputs of sense amplifier is enlarged and shown in Fig. 6(b). Under the worst scenario, there is $18\mathrm{mV}$ difference, which can be successfully sensed out at $500\mathrm{MHz}$ operating frequency.

The process variation induced resistance shifting is another severe issue in RRAM design. We conducted an evaluation under the extreme condition by increasing the resistances of data cells and decreasing the resistance of reference cells. Similar to the previous simulations, the critical data pattern was applied. Fig. 7(a) shows sense amplifier inputs "in" and "ref" when varying RRAM resistance 5% to 30% from its mean values, at a step of 5%. The corresponding output of the sense amplifier is shown in Fig. 7(b). The simulation results show that the proposed scheme can tolerate up to 20% variation of RRAM resistance and successfully read out the stored data. When the variation is more than 20%, the readout data is errant under the most critical condition.

The result in Fig. 7 demonstrates our proposed sense amplifier design has a sensing margin of $\sim 10\mathrm{mV}$. When the signal difference between the two inputs is smaller than $10\mathrm{mV}$, *i.e.*, the two fail cases in which RRAM resistance deviates 25% or 30% from its designed value, sensing operation fails. A better noise cancellation circuitry can further improve the sensing margin with the price of area overhead.

At last, we compared the various sense amplifier designs [16] used in RRAM crossbar array in terms of sense speed and design area. The summarized results in TABLE I show that the latch comparator in this work over performs the other sense amplifier designs. The *voltage divider* and *TIA* schemes generate the reference signals by using reference resistor, which cost a large area in modern technology. The *sigma delta* designs use an internal capacitor for integrating generated current that significantly slows down the sensing speed and increases design area. Besides the latch comparator, the proposed design also benefit from the expanded crossbar array structure, which can better trace the impact of data pattern with negligible area overhead.

TABLE I

COMPARISON OF VARIOUS SENSE AMPLIFIER DESIGNS

	Sense Speed	Area (normalized)
Latch Comparator (this work)	<4 ns	1.00
Voltage Divider [16]	<50 ns	1.78
TIA-based [16]	<100 ns	1.78
Sigma Delta w/ Buffer [16]	<10 us	7.29
Sigma Delta w/o Buffer [16]	<50 us	8.51

V. CONCLUSION

In this paper, we propose the novel peripheral circuit, including the compact decoder, the expanded crossbar array, and the latch comparator based sense amplifier, for the RRAM-based LUT. Benefiting from high density of the RRAM array, the simple peripheral circuit, and the bit addressability of new decoder, the proposed LUT can dramatically reduce design area, lower access latency, and provide dynamic reconfiguration. The advantages of the proposed design becomes even more obvious when increasing LUT input number.

REFERENCES

[1] Altera, "FPGA Architecture White Paper," 2006.
[2] ——, "MAX II Architecture," 2008.
[3] Xilinx, "Vertex 7 Series FPGAs Overview," 2011.
[4] S. Paul and S. Bhunia, "Computing with nanoscale memory: Model and architecture," in *IEEE/ACM NANOARCH*, 2009, pp. 1–6.
[5] H. Yan, et al., "Programmable nanowire circuits for nanoprocessors," *Nature*, vol. 470, no. 7333, pp. 240–244, 2011.
[6] I. G. Baek, et al., "Multi-layer cross-point binary oxide resistive memory (OxRRAM) for post-NAND storage application," in *IEEE IEDM*, 2005, pp. 750–753.
[7] P. Chow, et al., "The design of a SRAM-based field-programmable gate array-Part II: Circuit design and layout," *IEEE TVLSI*, vol. 7, no. 3, pp. 321–330, 1999.
[8] D. Lewis, et al., "Distributed memory in field-programmable gate array integrated circuit devices," Feb. 2 2010, US Patent 7,656,191.
[9] R. Waser, et al., "Redox-Based Resistive Switching Memories–Nanoionic Mechanisms, Prospects, and Challenges," *Advanced Materials*, vol. 21, no. 25-26, pp. 2632–2663, 2009.
[10] B.J. Choi, et al., "Purely Electronic Switching with High Uniformity, Resistance Tunability, and Good Retention in Pt-Dispersed SiO2 Thin Films for ReRAM," *Advanced Materials*, 2011.
[11] C.H. Ho, et al., "9nm Half-Pitch Functional Resistive Memory Cell with$<1\mu$A Programming Current Using Thermally Oxidized Sub-Stoichiometric WOx Film," in *IEEE IEDM*, 2010, pp. 436–439.
[12] J.P. Strachan, et al., "The switching location of a bipolar memristor: chemical, thermal and structural mapping," *Nanotechnology*, vol. 22, p. 254015, 2011.
[13] Y. Chen, et al., "3D-HIM: A 3D High-density Interleaved Memory for Bipolar RRAM Design," in *IEEE/ACM NANOARCH*, 2011, pp. 59–64.
[14] S.S. Sheu, et al., "A 4Mb embedded SLC resistive-RAM macro with 7.2 ns read-write random-access time and 160ns MLC-access capability," in *IEEE ISSCC*, 2011, pp. 200–202.
[15] B. Goll and H. Zimmermann, "A 0.12μm CMOS Comparator Requiring 0.5 V at 600MHz and 1.5 V at 6GHz," in *ISSCC*, 2007, pp. 316–605.
[16] M.S. Qureshi, et al., "CMOS Interface Circuits for Reading and Writing Memristor Crossbar Array," in *IEEE ISCAS*, 2011, pp. 2954–2957.