Fully BEOL-Compatible Switch Boxes Using RRAMs and Thin Film Transistors for Reconfigurable and Secure ICs

Aaron Ruen
Department of Electrical Engineering & Computer Science
University of Cincinnati
Cincinnati, Ohio 45221
ruenan@mail.uc.edu

Abhijeet Barua
Department of Electrical Engineering & Computer Science
University of Cincinnati
Cincinnati, Ohio 45221
baruaat@mail.uc.edu

Rashmi Jha
Department of Electrical Engineering & Computer Science
University of Cincinnati
Cincinnati, Ohio 45221
jhari@ucmail.uc.edu

John M. Emmert
Department of Electrical Engineering & Computer Science
University of Cincinnati
Cincinnati, Ohio 45221
john.emmert@uc.edu

***Abstract*—In this paper, we report Resistive Random Access Memory (RRAM) and Thin Film Transistors (TFT) based Switchboxes (SB) for Integrated Circuits (IC) design obfuscation and secure manufacturing. Our work shows that RRAM-TFT provide promising route for implementing SBs without compromising area in CMOS front end. Finally, we demonstrate its application for design obfuscation and reconfiguration of asynchronous logic cells.**

Keywords—Switchbox, reconfiguration, RRAM, Obfuscation, Security

I. INTRODUCTION

In recent years, there has been an increase in concerns surrounding hardware trust and security due to the globalization of chip fabrication [1]. Most foundries are located outside of the United States. There are various places in the supply chain where chip designs can be reverse engineered, duplicated, counterfeited, stolen, and otherwise compromised. Design obfuscation has become a necessary countermeasure to this sort of malicious activity. Design obfuscation can prevent reverse engineering in two ways. First, it can prevent the malicious agent from learning and reproducing the chip's design and function when they have obtained a physical copy of the chip. Second, it can prevent an untrusted foundry from ascertaining the function and proper operation of the chip even when they have the actual chip design. A common form of design obfuscation is via insertion of switchbox (SB). SBs can be programmed by the designer after the fabrication in untrusted foundry and allow for the routing of signals within a circuit to a desired location known to the designer and end user, but not to an untrusted foundry. A common way that this is implemented is with pass transistors with SRAM memory [2][5]. However, SRAMs are known to suffer from issues such as scalability, increased power consumption, and are expensive to fabricate. Additionally, these are volatile-memory devices which means reconfiguration bits needs to be stored in an on-chip non-volatile memory in case of power-loss. Resistive Random Access Memory (RRAM) devices have emerged as a popular option for replacing or augmenting SRAM-based SBs in hardware [4][5][6]. RRAM devices are scalable, reprogrammable, and non-volatile, making them an attractive option since they do not need to be reprogrammed every time power is cycled. RRAM devices and associated programming circuitry have an application in SBs designs for FPGA-based architectures. However, prior research have explored RRAMs with CMOS-Field Effect Transistors (FETs) where CMOS-FETs were fabricated in Front End of Line (FEOL) that consumes the expensive area on the chip in FEOL [2][5]. Our approach uses RRAM devices in conjunction with thin film transistors (TFTs) that can be completely fabricated into Back-End-of-Line (BEOL). This approach will allow for split-manufacturing with ease or additive manufacturing where RRAM-TFT-based SBs can be integrated on pre-fabricated CMOS dies at low-volumes in trusted foundries. Finally, we demonstrate application of RRAM-TFT based SBs for asynchronous reprogrammable gates to provide a scalable approach to achieving secure asynchronous FPGA designs.

II. OVERVIEW AND PAPER OUTLINE

A. Threat Model and Mitigation

Our proposed approach mitigates side-channel attacks by providing asynchronous FPGA design, and design obfuscation of asynchronous FPGA using SBs to protect against intellectual property (IP) reverse engineering, overproduction, and counterfeiting. The asynchronous logic designs have been proposed to provide solutions against Side-Channel Attacks (SCA) [1]. Addition of SBs will allow for reconfigurability as well as design obfuscation. Fig.1 shows the supply chain for asynchronous IC and how insertion of SBs can make it more secured against overproduction and counterfeiting. However, the current switchbox designs rely on SRAM technology and take up space in the logic layers. By constructing this switchbox as an additive layer, several major advantages can be achieved. The first advantage of this technique is that it provides for

flexibility of manufacturing. The fabrication processes can all be done in BEOL processes allowing the addition of another layer and allowing more space in other layers. Since the proper programming bits for these SBs is only known by the designer or legitimate customers, it obfuscates the design from any malicious agents or untrusted foundries that have obtained either a physical copy of the chip or the chip design, respectively. Finally, this method allows for reconfigurable designs.

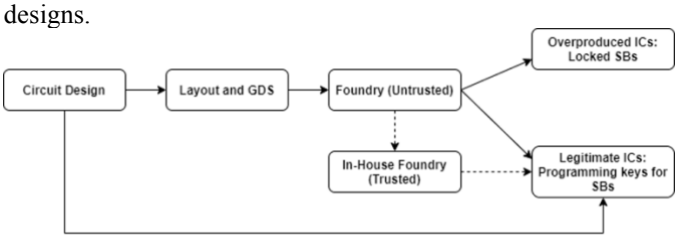

Fig. 1. Asynchronous IC supply chain and mitigation using RRAM-TFT SBs.

B. Switchbox

Switchboxes are a common technology used to route signals based on programming bits to provide reconfigurability and obfuscation to designs [2][3][6]. Fig. 2 shows a SB where by programing appropriate bits (0/1) on the gates of pass transistors, any input i, iϵI can be connected to any output o, oϵO. Programming bits are usually stored in SRAMs in a typical SB. For obfuscation purpose, only one (i,o) combination is correct which is known only to the designer. By careful insertion on SBs, attack complexity for unprogrammed SBs by attackers could be made very high (2^n, where n is the number of programming bits in SB) [3]. Additionally, under scenarios where SBs are integrated monolithically in trusted foundry (like proposed in this work), attacker will not have access to the full hardware to implement the SB which makes design very secure against attacks.

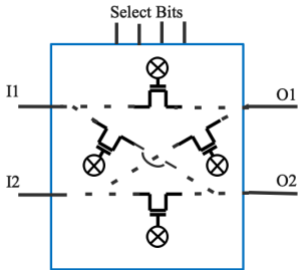

Fig. 2. A generic SB with pass transistors, and programmed bits stored in SRAMs.

We propose to design these SBs using RRAM-TFT where RRAM is used as non-volatile bit storage element and TFT is used as programming transistors for RRAM. Due to low thermal budget of RRAM and TFT, this entire circuitry can be integrated in BEOL or additively on CMOS dies. To make the SB as described and verify its operation; a TFT model, a RRAM model, and simulation platforms are required. Remainder of the paper is organized as: section II discusses the RRAM devices including their fabrication and modelling; section III discusses the TFT experimental results and modeling; section IV discusses the design of the RRAM-TFT SB and their simulation; section V discusses the modelling and simulation of the devices in a SPICE simulator. Lastly, Section IV discusses the results.

II. RRAM Device Fabrication, Testing, and Modeling

The RRAM devices were fabricated on Si/SiO2 substrate consisting of a Bottom Electrode (BE) made of Ti/Ru, 8nm HfO2, and 8 nm Zr Interfacial Layer, and 40nm of W as top electrode (TE). These fabricated and characterized RRAM devices had an area of 100um x 100um. The switching characteristics of these devices are shown on Fig. 3. and the fabricated architecture shown in Fig. 4. The devices required a Set voltage of approximately 0.75V and a RESET voltage of approximately -2V. These programming voltages required a maximum switching current of 1mA, which meant that the TFT devices had to be able to provide this current. The experimental data for the RRAM devices were collected on a Keithley 4200-SCS fitted with a 4225 PMU and a Cascade Microtech MPS150 probe station.

To design and develop SBs, RRAM switching models needed to be created to fit the experimental data. The model was fit to a complete switching cycle starting from a Formed state to show fit to experimental data. Since the model starting point is the Form state, it is appropriate to show device response to a post-Forming Reset operation and compare the state change to a post-Set Reset operation. Since the Set state and Formed state are significantly spread in Resistance, a split between achieved Reset state should be demonstrated with equivalent voltage. Noted from the experimental data, tested devices tend to Form very close 200Ω. From this stage, a significant Reset voltage is needed to retract the filament enough for large HRS to LRS switching ratio, around -2V, producing a 20kΩ state. A

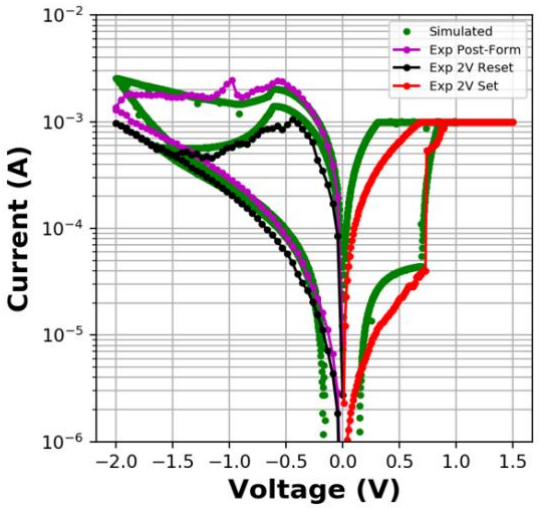

Fig. 3. RRAM model simulated data fitted to the experimentally measured data for switching curve.

Identify applicable funding agency here. If none, delete this text box.

Set operation completed after Reset, using the max current of 1mA, produced a 500Ω LRS. Finally, -2V was again used to Reset, producing a 25kΩ state. Experimental curve and fit to experimental curve using RRAM model is shown in Fig. 3.

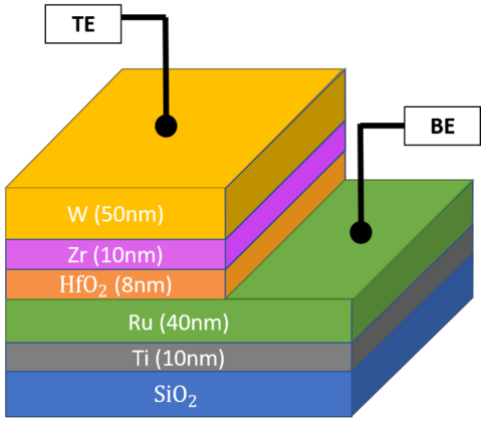

Fig. 4. RRAM fabricated device architecture.

III. Thin-Film Transistors Fabrication, Testing, and Modeling

There are several candidates for TFT applications, however, Indium Gallium Zinc Oxide (IGZO) channel-based TFT have gathered much attention. IGZO is a stable three-cation quaternary compound, and wide bandgap unipolar semiconductor (3.25 to 3.3eV) that demonstrates reduced leakage due to the wide bandgap, high mobility, low process complexity, low thermal budgets, good flexibility, and stable threshold voltages. Indium (In) increases oxygen vacancies in the overall amorphous microstructure which leads to an abundance of unbound metal atoms that leave more electrons free to increase the carrier density. The mobility is enhanced as the s orbital is isotropic and larger. This leads to better overlap between successive orbitals and highly mobile ns transport. However, binary oxides of In may be unsafe above critical cytotoxicity levels, because among In, Ga, and Zn, it is the only highly toxic material and therefore a ternary combination makes more sense. Ga is known to suppress mobility to semiconductor levels as it bonds O strongly than In or Zn, which creates a stable metal-O-metal structure. Zn has a strong tendency to crystallize along with mobility enhancement and exhibits small effective mass which decreases the effective density of states and tends to suppress the electron concentration to semiconductor levels. Ga, In, and especially Zn may individually be easily etched by wet methods, but this can be tackled by a ternary combination. The amorphous nature of the material requires low-temperature processing that enables incorporation of standard deposition methods compatible with CMOS BEOL processes.

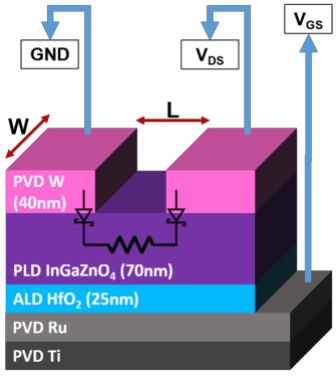

Fig. 5. TFT fabricated device architecture.

A bottom gate top contact inverted staggered architecture was used for the fabrication of IGZO TFTs on a 4-inch p - Si/SiO2 wafer substrate, shown in Fig. 5. First, 10 nm Ti adhesion promoter and 70 nm Ru were deposited in situ without any vacuum break at room temperature (RT) through radio frequency (RF) magnetron sputter deposition in an Ar environment at a flow rate of 12 sccm, a process pressure of 5.44 to 5.47 mTorr, and RF power of 100 W to form the global bottom gate. Second, a high-k HfO2 gate dielectric was chosen to achieve a high gate capacitance and enable the accumulation of a higher charge density in the channel at lower voltages with simultaneous suppression of gate leakage. Amorphous HfO2 film with short-range order was deposited in a Cambridge Nanotech Fiji F200 atomic layer deposition (ALD) system at 250 °C. Tetrakis(dimethylamido)hafnium (IV) was used as the hafnium source and a remote radio frequency O2 plasma at 300 W was used as the oxygen source yielding a growth rate of 1Å/cycle. Ar was used as a precursor carrier gas and plasma purge gas. Third, amorphous IGZO films were deposited in a

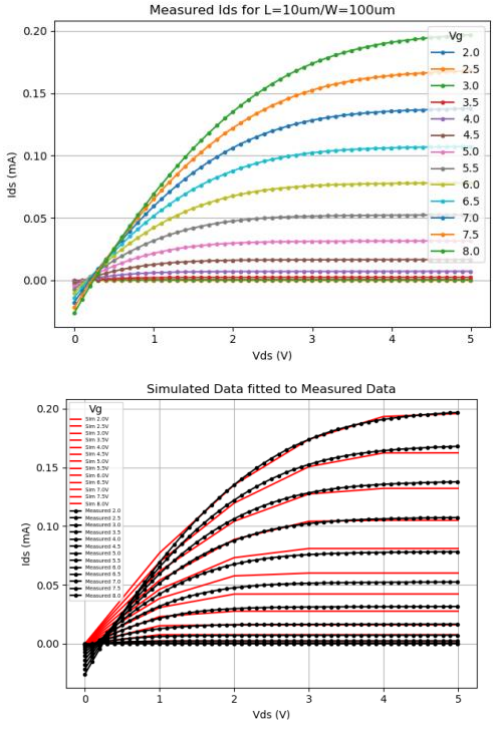

Fig. 6: (a) Measured Ids vs. Vds for various Vg values for the W/L=100um/10um TFT device, (b) Simulated data using the SPICE nmos model fitted to the measured device data. Based on RRAM switching current, The TFT should be able to provide a target 1mA of current.

Neocera Pioneer 180 pulsed laser deposition (PLD) system with a KrF excimer laser (Lambda Physik COMPex Pro 110, λ = 248 nm, 10 ns pulse duration), laser energy density of 2.6 J/cm2 , laser repetition rate of 30 Hz, deposition temperature of 25 °C, oxygen partial pressure of 25 mTorr, and substrate-to-target distance of 9.5 cm. The target was a 50 mm diameter by 6 mm thick sintered ceramic disk of In2O3 :Ga2O3 :ZnO with a 1:1:5 molar ratio. IGZO post-deposition anneal was performed at 200°C to recover device properties. The S/D contacts on IGZO channel was achieved using W (Tungsten). Active regions with channel lengths (L) of 10/20/50/100 µm and three common channel widths (W) of 50/100/150 µm were patterned with a positive etch mask in an EVG 620 Mask Aligner. No pre- or post-fabrication anneal was performed on the wafer such that virgin characteristics were preserved and the maximum process temperature used was only in ALD. The device regions were cleaved for further electrical characterization through a Keithley 4200A Semiconductor Characterization System (SCS) with a Cascade Microtech MPS150 Probe Station at room temperature (RT) of 298.15 K under ambient environment, and a Lakeshore Cryotronics PS-100 Cryogenic Probe Station with TTPX Model 336 Controller, under vacuum environment.

The voltage bias applied to the gate terminal of the devices was increased from 2V to 8V in 0.5V steps. For each unique voltage applied to the gate, the voltage bias applied to the drain terminal of the devices was swept from 0V to 5V in 0.1V steps while the gate voltage was held constant for the duration of the sweep. Likewise, a similar experiment was done where various voltage biases were applied to the drain terminal while the bias applied to the gate was swept. The Ids vs. Vds characteristic is shown in Fig. 6(a) at various Vgs.

Appropriate TFT model is necessary for accurately simulating circuit behavior. Fitting was obtained by changing the default values in the SPICE NMOS model for the threshold voltage (Vto), Transconductance parameter (Kp), Channel-length modulation (Lambda), and the bulk threshold parameter (Gamma). The values of 2.4, 1.55e-6, 0.01, and 0.75 for Vto, Kp, Lambda, and Gamma, respectively were chosen to obtain the fitted data shown in Fig. 6(b). From these plots, a Vto of approximately 2.4V was achieved.

As can be observed in Fig. 6(b), excellent fitting of the model to the experimentally measured IGZO TFT curve can be obtained. Next, since RRAM needed 1mA of switching current, it was necessary to find TFT dimensions and bias voltages that could provide this. To obtain this, the normalized current was calculated by dividing the current by the ratio of the Width to the Length of the channel. The maximum measured current was approximately 0.2mA. Since 1mA of current is required to properly program the RRAM devices, different TFT device dimensions were required. Acceptable device dimensions were acquired by performing a design space exploration were combinations of channel length and channel width were explored with a drain-to-source current of at least 1mA at a drain voltage less than or equal to 5V and a gate voltage less than or equal to 8V was the goal. Fig. 7 shows that for lower gate voltages, such as 2V, there are no reasonable device dimensions that would be able to provide the required current for Vds less than or equal to 5V. At a Vgs of 5V, there are several dimensions that are projected to be able to provide the necessary 1mA, however the W is quite large. Furthermore, above a Vgs of 7V, the devices will be able to provide 1mA, with reasonable device widths.

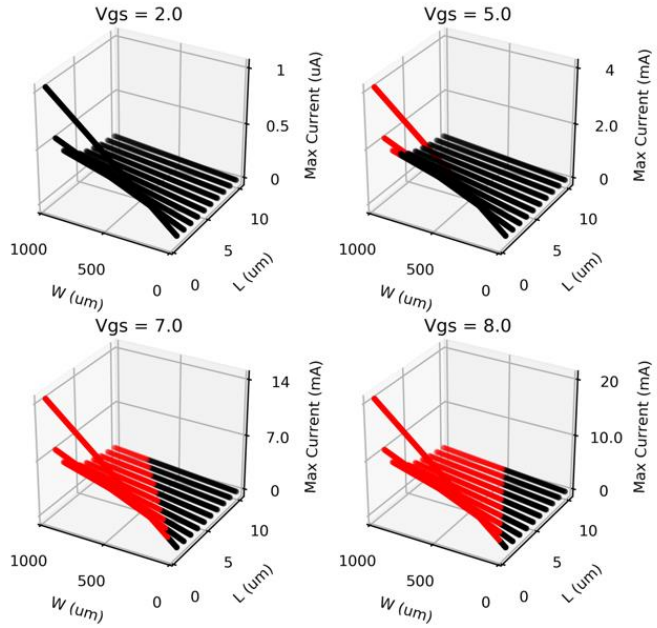

Fig. 7. Design Space Exploration for the maximum current values able to provide for any Vds (0V to 5V) and Vgs (0 to 8V) for various W and L dimensions.

IV. Design of RRAM-TFT SB

Fig. 8 shows RRAM-TFT SB design. The RRAM and TFT devices are shown in red that can be fabricated in BEOL, the pass transistor is shown in blue that is Si-NMOS transistor that needs to be fabricated in FEOL. Compared to Si-CMOS FETs, TFTs offer lower drive current, higher ON resistance, higher VT, and slower switching speeds. Therefore, if pass transistors are designed using TFTs then there is a possible speed penalty which is undesired. Therefore, in our design pass transistors are still designed using Si-CMOS FETs. However, TFTs can be used to implement programming circuits for RRAM as SBs will

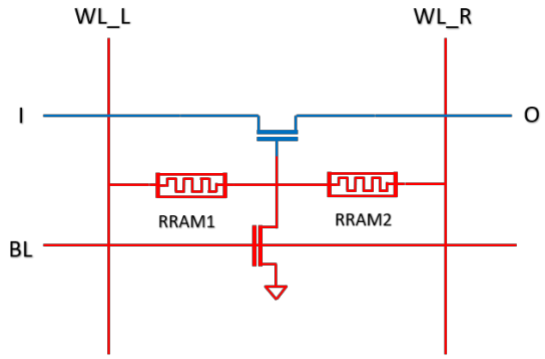

Fig. 8: RRAM-TFT SB with pass transistor (blue) and programming transistor (red) with selection lines.

be reconfigured off-line and infrequently and speed and high-ON resistance of TFT do not impose limitations. However, TFTs need to be carefully designed to handle the switching current requirements of RRAM. In our case, the maximum current required by RRAM is 1mA, therefore, based on our design space exploration data, TFT of W/L= 100um/1um with Vgs = 7V ; Vds = 2V was appropriate. The devices are programmed to turn ON (bit 1) the pass transistor by putting RRAM1 in LRS and RRAM2 in HRS. The devices are programmed to turn OFF (bit 0) the pass transistor by putting RRAM1 in HRS and RRAM2 in LRS. RRAM1 is placed into LRS by applying a voltage of 2V on WL_L, 7V on the BL, and 0V on WL_R. This also places RRAM2 in HRS. To place RRAM1 in HRS and RRAM2 to LRS; apply 2V to WL_L, 7V on the BL, and 0V on WL_R. After programming, during normal operation, to achieve the desired bias on the pass transistors, a small read voltage of less than 0.5V needs to be applied on BL_L with BL_R grounded. Various programming circuitries have been proposed for RRAM applications and commonly used in purely memory applications is the 1T1R configuration. Another programming circuitry has been discussed in [1] as usable for a switchbox in a synchronous FPGA design. This programming circuitry uses left and right bitlines to program two RRAM into opposite states. This programming circuitry is not tied to the synchronous FPGA and can be scalable. Therefore, we make use of this programming scheme with the condition that the programming transistors in the design have been replaced TFTs.

V. Design and Simulation of Reconfigurable Asychronous Logic with RRAM-TFT SB

The asynchromous logic, simulated in our work, is shown in Fig. 9. When M=1 (Mb=0); the THx2 [1] this asynchronous circuit will operate in the TH12 mode. The TH12 operation is an asynchronous OR gate; when M=0 (Mb=1); circuit operates will operate in the TH22 mode. The TH12 mode operates as an asynchronous AND gate. In this case, M and Mb will work as pass transistors while programming bits will be provided by RRAM-TFT SB, discussed in section IV.

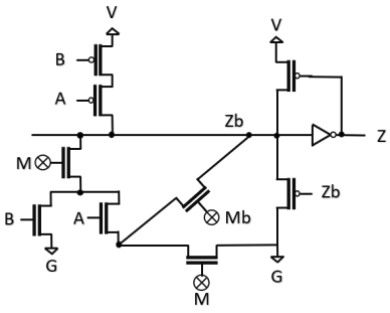

Fig. 9: THx2 cell as described in [2]

To demonstrate this, the simulations were performed using the following process: First, a circuit netlist was created using ltspice. Second, this netlist is read into Python using the PyLTSpice package. Third, the batch of simulations to be run is set up in python. These simulations hold the RRAM device information in Python and provide current and voltage biases from ltspice to the RRAM python model and return necessary information to be added to the ltspice netlist. Essentially, the simulation platform generates time-based lookup tables for the RRAM nodes. All non-RRAM components are simulated purely in ltspice, but Python acts as the simulation interface and provides the method by which the RRAM model is integrated into the ltspice simulation. The results are shown in Fig. 10 where the gates could be successfully reconfigured between TH12 and TH22.

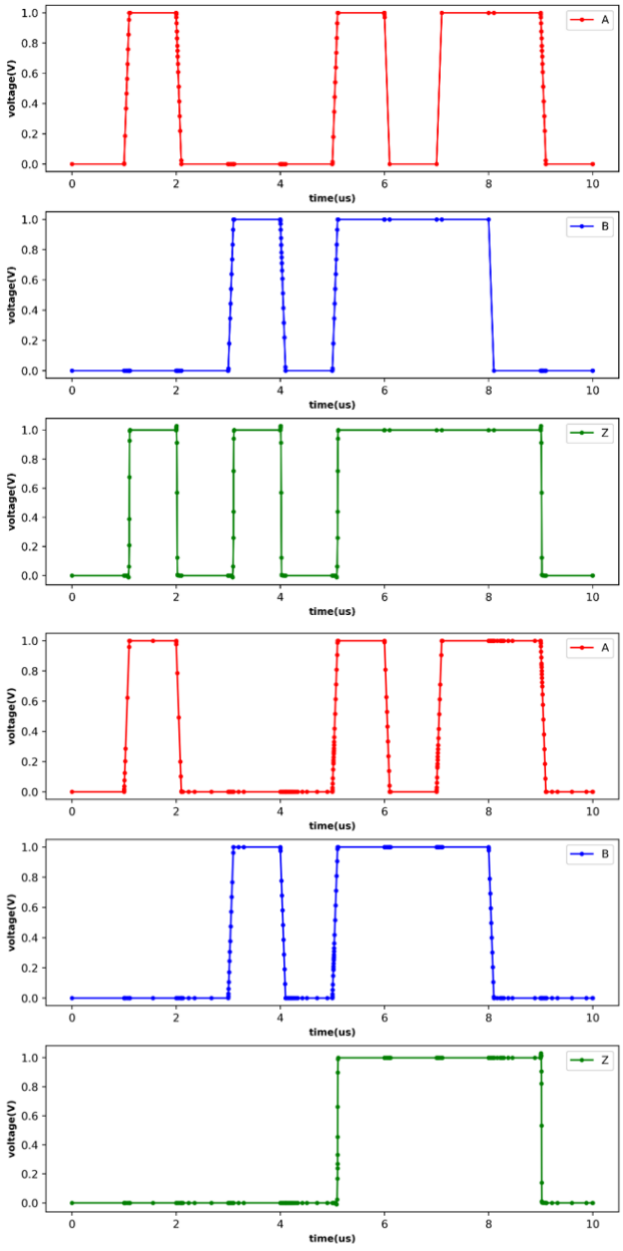

Fig. 10: Using the RRAM SB as the inputs to the control signals M, (a) THx2 operating as TH12 asynchronous OR gate. (b) THx2 operating as TH22 asynchronous AND gate.

VI. Conclusion

In conclusions, we developed a novel RRAM-TFT SB for design obfuscation and demonstrated its application for asynchronous reconfigurable circuit. The major advantage is RRAM-TFT is that the complete SB can be integrated in BEOL or additively which provides promising possibilities of secure manufacturing without compromising area in FEOL. SB sizes can be significantly reduced by reducing the RRAM switching currents which will be the focus of our future work.

Acknowledgment

This work is funded by National Science Foundation REU supplements to award no. CCF-1718428 and AFRL/ECI award no. IMPACTS and TAME. Authors would like to thank Mr. Nicholas Olexa at the University of Cincinnati for developing the RRAM model.

References

[1] *Emmert, J. M. (n.d.). An Asynchronous FPGA TH x 2 Programmable Cell for Mitigating Side-Channel Attacks. 3–6. https://doi.org/10.1109/MWSCAS48704.2020.9184563*

[2] *Gaillardon, P. E., Tang, X., Sandrini, J., Thammasack, M., Omam, S. R., Sacchetto, D., Leblebici, Y., & De Micheli, G. (2015). A ultra-low-power FPGA based on monolithically integrated RRAMs. Proceedings -Design, Automation and Test in Europe, DATE, 2015-April, 1203–1208. https://doi.org/10.7873/date.2015.1112*

[3] *Guo, Z., Tehranipoor, M. M., & Forte, D. (2017). Permutation-Based Obfuscation. In D. Forte, S. Bhunia, & M. M. Tehranipoor (Eds.), Hardware Protection through Obfuscation (pp. 103–133). Springer International Publishing. https://doi.org/10.1007/978-3-319-49019-9_5*

[4] *Nguyen, T. H., Barua, A., Bailey, T., Rush, A., Kosel, P., Leedy, K., & Jha, R. (2018). Reflection coefficient of HfO2-based RRAM in different resistance states. Applied Physics Letters, 113(19), 192101. https://doi.org/10.1063/1.5034118*

[5] *Palma, G., Vianello, E., Thomas, O., Suri, M., Onkaraiah, S., Toffoli, A., Carabasse, C., Bernard, M., Roule, A., Pirrotta, O., Molas, G., & De Salvo, B. (2014). Interface Engineering of Ag-GeS2-based conductive bridge RAM for reconfigurable logic applications. IEEE Transactions on Electron Devices, 61(3), 793–800. https://doi.org/10.1109/TED.2014.2301694*

[6] *Wu, T. F., Ganesan, K., Hu, Y. A., Wong, H. S. P., Wong, S., & Mitra, S. (2016). TPAD: Hardware trojan prevention and detection for trusted integrated circuits. IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems, 35(4), 521–534. https://doi.org/10.1109/TCAD.2015.24743*