Fine-Grained 3D Reconfigurable Computing Fabric with RRAM

Mingyu Li, Jiajun Shi, Sachin Bhat, Csaba Andras Moritz
Electrical and Computer Engineering, Univ. of Massachusetts, Amherst, USA

***Abstract*—Non-volatile 3D FPGA research to date utilizes layer-by-layer stacking of 2D CMOS / RRAM circuits. On the other hand, vertically-composed 3D FPGA that integrates CMOS and RRAM circuits has eluded us, owing to the difficult requirement of highly customized regional doping and material insertion in 3D to build and route complementary p- and n-type transistors as well as resistive switches. In the layer-by-layer non-volatile 3D FPGA, the connectivity between the monolithically stacked RRAMs and underlying CMOS circuits is likely to be limited and lead to large parasitic RCs. In this paper, we propose a fine-grained 3D reconfigurable computing fabric concept. It implements CMOS / RRAM hybrid circuits within the pre-doped vertical nanowire template. Transistors and resistive switches can be integrated with fine granularity, which reduces the routing overhead between RRAM and CMOS circuits and increases the density. We estimate the density benefit of the proposed fabric to be 27X relative to the monolithic 3D FPGA with stacked RRAMs. Estimated Elmore delays are improved by 5.4X and 2.2X for configuration and normal operation, respectively.**

Keywords—fine-grained 3D reconfigurable computing, RRAM

I. INTRODUCTION

Field Programmable Gate Arrays (FPGAs) can provide flexible computing platforms but are fundamentally inefficient relative to custom Application Specific Integrated Circuits (ASICs), which is largely also due to the large interconnection and configuration overhead [1]. Research directions to date attempt to address these issues by combining the emerging 3D integration and Resistive RAM (RRAM) technologies [2]-[5] as 3D integration is a promising option to overcome the interconnection bottleneck, and RRAM can greatly improve the configuration memory density. However, it is still considered infeasible to have a vertically-composed 3D reconfigurable computing fabric which at the same time incorporates CMOS and RRAM. One major issue is that such a technology would require highly customized regional doping and material insertion in 3D, to form and route the CMOS and RRAM circuits. Because of these difficult requirements, the state-of-art research focuses on die-to-die or layer-by-layer approaches, with each layer only consisting either CMOS or RRAM circuits in 2D [2]-[5]. RRAM arrays are usually monolithically stacked on top of the CMOS circuits, which greatly improves the density of the configuration memories. However, the wiring between RRAM and CMOS circuits can be difficult and inefficient. Figure I(A)-(D) provides an intuition on the issues that 3D RRAM-based FPGA directions could encounter.

In this paper, we propose a fine-grained 3D reconfigurable computing fabric concept. All circuits are realized in and across uniform vertical nanowires, which are pre-doped in horizontal stripes. We place and connect transistors on these nanowires either in series (on one nanowire) or in parallel (across multiple nanowires) to build CMOS circuits; RRAM cells storing configured bits are placed directly next to the transistors they control, which reduces wiring cost and achieves high density. This fabric does not add manufacturing requirement to the ones used for various Skybridge directions, which implement fine-grained 3D ICs [6]-[8].

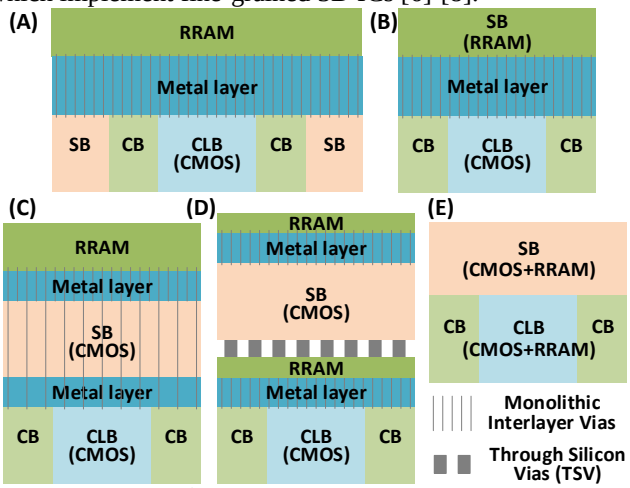

Figure I. Side views of other 3D RRAM FPGA directions in (A-D) vs. our direction in (E) (CLB: Clustered Logic Box, CB: Connection Box, SB: Switch Box): A). RRAMs on 2D CMOS [2], CMOS density not improved; B). RRAMs on 2D CMOS [3], SB implemented only with RRAMs C). RRAMs on the topmost layer above the underlying 3D CMOS [4], connecting RRAM and bottom CMOS layers is difficult and inefficient; D). RRAMs stacked on each of the layers in 3D CMOS [5]; top CMOS circuits have to connect through TSVs (as opposed to be monolithically stacked on top of RRAMs) due to the incompatible process; in (A)-(D), RRAMs are monolithically stacked, connecting RRAMs and CMOS is through all metal layers and vias, leading to large parasitic RCs; E). our proposed fabric, each layer contains CMOS / RRAM hybrid circuits, vertical routing fabric components provide good connectivity between layers

II. FABRIC COMPONENTS AND BASIC FUNCTION BLOCKS

The afore-mentioned fine-grained 3D FPGA incorporating CMOS and RRAM technologies is enabled by the following fabric components.

The RRAM cells in our fabric, as opposed to being stacked on the 2D CMOS circuits in an array, are distributed and with fine granularity integrated with the other fabric components, including transistors and routing elements. As Figure II shows, the resistive switch is built with layers of material deposition surrounding a nanowire, which does not depart from the manufacturing requirement of other components in this fabric. The inner electrode forms an Ohmic contact with the heavily doped nanowire, allowing us to access the inner electrode through the nanowire. The outer layer forms the other electrode. The two electrodes are slightly asymmetric in size; a device with similar asymmetric structure has been demonstrated experimentally [9]. The resistive switch has similar feature size with 16nm transistors; devices with the

feature size of 10x10nm^2 have been experimentally validated [10]. Our RRAM devices has Ti electrodes and TiO_2 switching layer [5]. Devices based on Al_2O_3 and HfO_2 switching layers have also been demonstrated [2] [10]. In each RRAM cell, we use a pair of complementary resistive switches which outputs a high or low voltage based on the configured information [2].

Our fabric also shares some fabric structures with the other Skybridge directions albeit differently integrated / utilized as shown in Figure II [7]. Vertical Gate-All-Around Junctionless transistors are the active devices in CMOS circuits. *Bridges* are horizontal wires connecting adjacent nanowires. Nanowires can be routing elements (*Routing Nanowires*) since they are heavily doped and silicided, and provide good conductivity. *Coaxial Routing structures* are metal layers along nanowires to add connectivity in the vertical direction. *Skybridge-Interlayer Connection (SB-ILC)* connects the p- and n-doped regions of a nanowire to form a signal path on a *Routing Nanowire*. It also connects the pull-up and pull-down networks to form a CMOS gate.

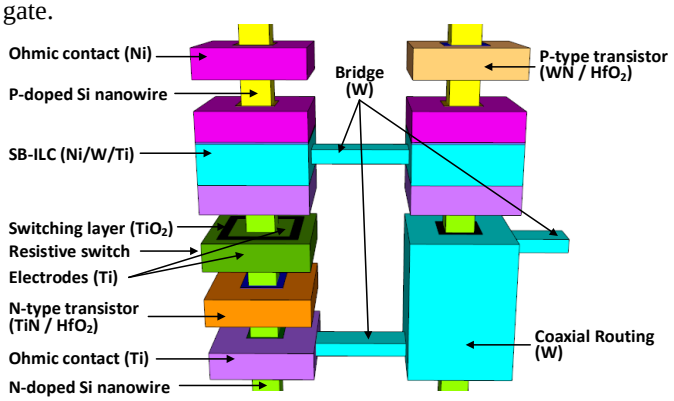

Figure II. Fabric components

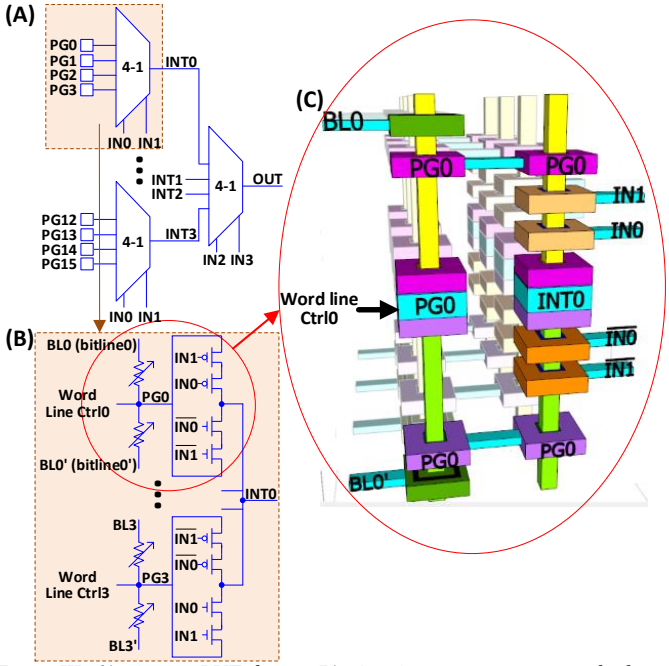

Figure III. A). 4-input LUT design; B). 4-to-1 transmission-gate multiplexer with configured inputs, and C). its layout with the circled part highlighted

The afore-mentioned novel fabric concepts allow us to build fine-grained 3D CMOS / RRAM hybrid reconfigurable circuits. Clustered Logic Boxes (CLBs) contain several Look Up Tables (LUTs) and flip-flops to emulate the functionality of combinational or sequential logic circuits. A 4-input LUT is shown in Figure III; it is implemented with a 16-to-1 multiplexer selecting one of the 16 configured bits based on the 4-bit input. Connection Box (CB) controls the connections between CLBs and routing tracks with pass transistors. Switch Box (SB) controls the routing directions where vertical and horizontal routing tracks intersect. Figure IV shows a part of SB that routes W0 from west to any of the three directions through pass transistors. Buffers are added when necessary.

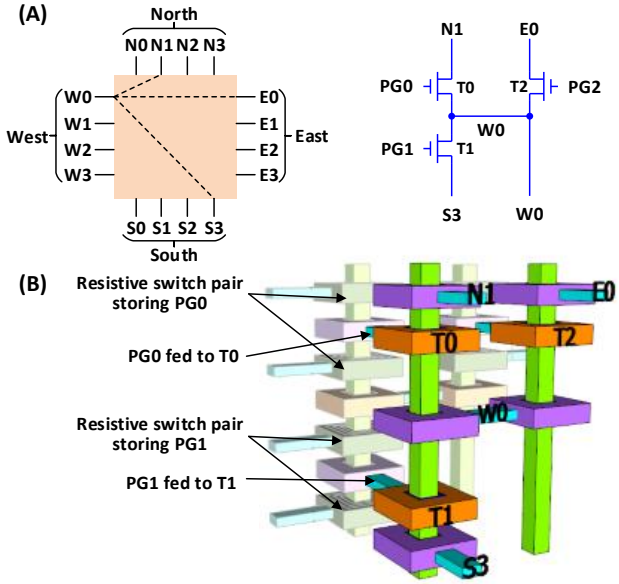

Figure IV. A). Wilton switch block and its schematic controlling connections of W0 (Wire 0 from the West direction), and B). its implementation with pass transistor network highlighted

III. Preliminary Evaluations and Conclusion

In this paper, a fine-grained 3D reconfigurable computing fabric is presented. As a part of our initial study, we compared the density of our fabric and the baseline 3D FPGA with stacked RRAM. We estimated the area footprint of one *tile* containing one CLB, one CB and one SB in both fabrics, and found that our fabric has 27X density benefits. Using the Elmore delay model, we have also estimated the configuration and operation delays of the circuits. Our fabric shows 5.4X benefits in configuration delay, which results from the reduced parasitic RCs on the path from an RRAM cell to CMOS layer due to the tightly integrated RRAM and CMOS circuits. The switching delay of a wire connecting two CLBs in neighboring tiles has been improved by 2.2X thanks to the higher density that reduces the wire length.

References

[1] V. George, Ph.D. dissertation, UC Berkeley, Berkeley, CA, 1999.
[2] Y. Y. Liauw, et al., *ISSCC*, pp. 406-408, 2012.
[3] P.-E. Gaillardon, et al., *DATE*, pp. 1203-1208, 2015.
[4] Y. Chen, et al., *ISLPED*, pp. 55-60, 2010.
[5] Y. C. Chen, et al., *FPL*, pp. 367-372, 2012.
[6] M. Li, et al., *ISVLSI*, pp. 403-408, 2016.
[7] M. Rahman, et al., [online]. Available: http://arxiv.org/abs/1404.0607
[8] J. Shi, et al., NANOARCH, pp. 169-174, 2015
[9] A. Williamson, et al., *Nanoscale*, vol. 5, pp. 7297-7303, 2013.
[10] B. Govoreanu, et al., *IEDM*, pp. 31.6.1-31.6.4, 2011.