Heterogeneous Technology Configurable Fabrics for Field-Programmable Co-design of CMOS and Spin-based Devices

Ronald F. DeMara, Arman Roohi, Ramtin Zand, and Steven D. Pyle Department of Electrical and Computer Engineering University of Central Florida Orlando FL, USA 32816-2362 November 7, 2017

Abstract—The architecture, operation, and characteristics of two post-CMOS reconfigurable fabrics are identified to realize energy-sparing and resilience features, while remaining feasible for near-term fabrication. First, Storage Cell Replacement Fabrics (SCRFs) provide a reconfigurable computing platform utilizing near-zero leakage Spin Hall Effect devices which replace SRAM bit-cells within Look-Up Tables (LUTs) and/or switch boxes to complement the advantages of MOS transistor-based multiplexer select trees. Second, Heterogeneous Technology Configurable Fabrics (HTCFs) are identified to extend reconfigurable computing platforms via a palette of CMOS, spinbased, or other emerging device technologies, such as various Magnetic Tunnel Junction (MTJ) and Domain Wall Motion devices. HTCFs are composed of a triad of Emerging Device Blocks, CMOS Logic Blocks, and Signal Conversion Blocks. This facilitates a novel architectural approach to reduce leakage energy, minimize communication occurrence and energy cost by eliminating unnecessary data transfer, and support auto-tuning for resilience. Furthermore, HTCFs enable new advantages of technology co-design which trades off alternative mappings between emerging devices and transistors at runtime by allowing dynamic remapping to adaptively leverage the intrinsic computing features of each device technology. Both SCRFs and HTCFs offer a platform for fine-grained Logic-In-Memory architectures and runtime adaptive hardware. SPICE simulations indicate 6% to 67% reduction in read energy, 21% reduction in reconfiguration energy, and 78% higher clock frequency versus alternative fabricated emerging device architectures, and a significant reduction in leakage compared to CMOS-based approaches.

Keywords— reconfigurable computing; FPGA; spintronics; post-CMOS architectures; energy-aware processor architecture; logicin-memory; resilient computing; MTJ; SHE device.

I. INTRODUCTION AND MOTIVATION

Similar to their ASIC counterparts, reconfigurable computing devices strive to surmount the growing technical challenges to improve their logic density, throughput performance, and power profiles. Thus with the geometrical and equivalent scaling trends guided by decades of International Technology Roadmap for Semiconductors (ITRS) projections nearing their end, new pathways towards these goals have been defined in ITRS 2.0 along with the IEEE International Roadmap for Devices and Systems (IRDS) initiative [1]. Two such technical thrusts identified for 2020 onward are *leveraging* beyond-CMOS devices (ITRS 2.0 theme 5) and utilizing heterogeneous components (ITRS 2.0 theme 4) to realize fundamentally new ways to compute. The perspective taken herein is that a reconfigurable computing paradigm can significantly advance both of these declared ITRS 2.0 themes.

Within the post Moore era, there are several motivations for pursuing novel reconfigurable fabrics of heterogeneous device technologies. Foremost, their one-time design and fabrication model minimizes the recurring engineering effort for post-CMOS devices, while amortizing development costs across multiple applications. Thus, reconfigurable fabrics may offer a more cost-effective approach to utilizing emerging devices. Additionally, post-CMOS ingrained field-programmable fabrics expand the accessibility of emerging devices to vast populations of circuit designers, including the majority of those who lack foundry access. Such a pre-fabrication approach with later fieldprogrammability minimizes the need for extensive post-CMOS circuit design, verification, and validation expertise. Fieldprogrammability also eliminates the computational demands, delays, and inaccuracies of simulation-based modeling associated with emerging devices. Instead, heterogeneous fabrics support rapid and direct realizations in hardware.

As a fundamentally different way to compute, the mapping of operations to device technologies remains fluid. Flexible mappings become possible not only during circuit synthesis, but also during execution-time. Thus when execution demands change, the architecture can adapt by utilizing a preferred device technology within its datapaths via reconfiguration of hardware components. This leverages the complementary characteristics of CMOS and emerging devices by increasing the flexibility in its binding of logic and memory roles to distinct device technologies. This is introduced herein as a post-CMOS era approach referred to as "technology co-design." Overall, the hypothesis is as follows: reconfigurable fabrics of heterogeneous CMOS and spin-based devices offer an orthogonal dimension of technology adaptation to balance throughput, energy consumption, and resilience beyond static emerging device architectures, fixed hybrid emerging/CMOS architectures, and CMOS-only reconfigurable platforms.

II. POST-CMOS RECONFIGURABLE FABRIC OPPORTUNITIES

Among promising spintronic devices, the ITRS Magnetism Roadmap identifies capable post-CMOS candidates of which Magnetic Tunnel Junctions (MTJs), Spin Hall Effect (SHE)

This research supported by the NSF Energy-Efficient Computing from Devices to Architectures (E2CDA) program #1739635 and SRC nCore Center Probabilistic Spin Logic for Low-Energy Boolean & Non-Boolean Computing.

Table I: Characteristics of emerg	ng device technol	logies. "🗸 " or	" indicates s	trength/limitation i	relative to	CMOS
-----------------------------------	-------------------	-----------------	---------------	----------------------	-------------	------

Attributes	Standby	Write	Density	Snood	Foundational Works		
Technology	Energy	Energy	Density	Speed	Text color indicates: NSF projects, academic efforts, commercial products		
CMOS	0	0	0	0	Xilinx & Intel provide Block RAMs and Multiply- Accumulate (MAC) blocks		
MTJ	$\checkmark\checkmark$		~		Configurable memory array; 3-input majority gate 50-200 mV swing; MRAM shipped by Everspin		
SHE	$\checkmark\checkmark$	-	✓	-	40μA non-volatile flip-flop;110x110x2 nm ³ ; 116x80x2.5 nm ³ 190 μA 0.4V STT-MRAM cell		
DMW	$\checkmark\checkmark$		 ✓ 		80 nm-width magnetic stripe as shift register; 2 ns, 34 μ m ² 8-bit adder		
NML	\checkmark	-	~		7–input majority gate; 17 μ m ² 1-bit perpendicular FA; 40 μ m ² 8-bit Discrete Cosine Transform		

enhanced switching devices, Domain Wall Motion (DWM) nanowires, and NanoMagnetic Logic (NML) are considered feasibly-implemented. In the case of STT-MTJs, they are currently commercially-available. Each technology's attributes relative to CMOS are listed in Table I, along with selected foundational works. Attributes complementary to CMOS are evident for spintronic devices, such as preferable static energy consumption, but a larger write energy than CMOS. Spintronic density is higher due to 3D vertical integration capability, although its switching speed is slower. Taking these characteristics into account, the key opportunities for spintronics in reconfigurable fabrics are summarized in Table II. Foremost, fabric flexibility allows a direct hardware realization which encapsulates device physics and expertise needed to design circuits using the targeted nanomagnetic devices. Applicationspecific hardware, including energy-aware designs, are able to leverage non-volatile SHE elements at medium and fine granularities via reconfiguration. Fabrics also allow in-situ localization of data stores and datapath re-construction at runtime based on changing execution demands and tradeoffs.

On the other hand in Table II, highly-scaled CMOS and emerging devices are susceptible to Process Variation (PV) effects. Reconfigurable computing paradigms have significant track records of mitigating PV challenges to maximize yield and resilience to hard faults. For instance, the fabric described in Section III builds upon MTJ memory reliability analysis methods and self-referencing reliability enhancement techniques [2]. Based on their relative attributes of both energy and resilience, SHE-enhanced switching devices are targeted to offer suitable replacements for SRAM cells in Look-Up Tables (LUTs) and switching blocks where configuration bit streams reside, as presented in Section III.

Meanwhile as depicted in Figure 1, the trend toward increasing heterogeneity within reconfigurable fabrics is wellestablished. Starting in the 1990s, various granularities of general-purpose reconfigurable logic blocks and dedicated function-specific computational units have been added to

Table II: Roles for Spintronics in Reconfigurable Computing.

Objective	Approach	Device	Role
Field- Programmability	Hardware realization w/o foundry access	All	Encapsulate device physics knowledge & design rules
Energy-Sparing	Data transfer reduced via local data stores	SHE	Fine-grained logic-in-memory w/ near-zero standby energy
Resiliency	Amorphous spares providing redundancy	SHE	PV adapt w/ reconfiguration & alpha-particle immunity
Adaptability	Datapath constructed based on demands	DML, NML	Leverage intrinsic switching / memory behavior of devices

fabrics. Combinational Logic Block (CLB) structures have resulted with increased computational functionality compared to homogeneous CLBs. Over the last ten years, reprogrammable fabrics have embedded an increasing number of special-purpose co-processing elements to handle complex floating-point computations, including DSP blocks, Multiplier-Accumulators (MACs), multi-bit block RAMs, and processor hardcores within commercially-available FPGAs.

It is proposed herein that fabric heterogeneity would be extended to the upper rightmost corner of Figure 1. It illustrates how emerging devices could advance technology-specific advantages, which is referred to as *technology heterogeneity*. As depicted with blue in Figure 1, the advantages of CMOS devices for rapid switching cooperate with spin-based devices that offer non-volatility, near-zero standby power, high integration density, and radiation-hardness. Realization of technology heterogeneity in a field-programmable fabric enables synthesistime co-design and dynamic run-time adaptability among a concise palette of devices, as depicted with ivory in Figure 1.

III. STORAGE CELL REPLACEMENT FABRICS (SCRFS)

As shown in Figure 2, Storage Cell Replacement Fabrics (SCRFs) provide a reconfigurable computing platform utilizing near-zero leakage spin-based devices to replace SRAM bit-cells within Look-Up Tables (LUTs) and within Switch Boxes of the routing network. LUTs implement a $2^m \times 1$ bit memory capable of realizing a Boolean logic function having *m* inputs. Currently, SRAM-based LUTs are a primary constituent for logic realization in most reconfigurable fabrics. However,



Figure 1: Escalation of field-programmable heterogeneity.



Figure 2: SCRF with Spin-based CLB and Switch Box configuration bit-cells (left) and design of SCRF Hybrid SHE and CMOS LUT (right) [6].

SRAM's drawbacks including high static power consumption, volatility, and restricted logic density have motivated exploration of alternative LUT designs by Zhao [3], Gaillardon [4], Suzuki [5], and others. They report measurements of fabricated non-volatile FPGAs with up to 81% power reduction over CMOS for representative applications [5]. This is achieved by leveraging the non-volatility feature of emerging resistive technologies, while attaining the related advantages identified in Figure 2 (left). The proposed fabric leverages the island-style network in its topology, as shown in Figure 2 (left). In this routing topology, Switching Blocks (SBs) connect horizontal and vertical routing tracks, while interconnection between CLBs and the routing network is via Connection Blocks (CBs) for local connections, as inspired by previous CMOS FPGAs.

Figure 2 (right) depicts the SHE-based 4-input Look-Up Table (LUT) we have developed as a building block for energyefficient non-volatile reconfigurable logic [6]. The read circuit select tree enables each storage cell according *A*, *B*, *C*, and *D* input address signals. Transmission Gate (TG) and Pass Transistor (PT) options were simulated indicating that TGs reduce delay and increase PV resilience at comparable power. In each SHE device, data is retained via resistive levels of Parallel (P) or Anti-Parallel (AP) spin configurations. To ascertain the P or AP resistive state, a Pre-Charge Sense Amplifier (PCSA) constructed with 7 MOS transistors compares the SHE's resistance to a reference MTJ value. Dimensions of the reference MTJ used for simulation were designed so the P configuration is between high resistance, R_{High} , and low resistance, R_{Low} , whereby $R_{P-reference MTJ} \cong$

Table III: 4-input LUT Read Op. (90nm CMOS, MTJ dimensions vary).

Attribute	Delay (ps)	Active Power (µW)	Energy (fJ)	EDP (fJ×ps)	Area (MTJ count)	Area (transistor count)
Zhao et al. [3]	88	13.4	1.179	103.8	32	154
Suzuki et al. [5]	81	7.58	6.140	497.3	36	74
Zand et al. [6]	94	4.3	0.404	38.0	17	112
SHE-based LUT	94.6	4.01	0.379	35.8	17	109

 $R_{AP-LUT MTJ} + R_{P-LUT MTJ})/2 + R_{HM}/2$ for Parallel (P), Anti-Parallel (AP), and Heavy Metal (HM) resistances, respectively. Results listed in Table III indicate 6% to 67% reduction in read energy, 21% reduction in reconfiguration energy, and 78% higher clock frequency versus alternative fabricated emerging devices reconfigurable architectures, as elaborated in [6]. Thus, these benefits are beyond the 81% energy reduction vs. CMOS.

IV. HETEROGENEOUS TECHNOLOGY CONFIG. FABRICS (HTCFs)

Beyond replacing SRAM cells within the fabric, Figure 3 depicts a longer-term vision toward Heterogeneous Technology Configurable Fabrics (HTCFs). HTCFs assimilate the complementary roles of a concise palette of spintronic and CMOS devices within a reconfigurable array. It is proposed herein that such heterogeneous fabrics be comprised by a triad of emerging device blocks, CMOS logic blocks, and signal conversion blocks. Emerging device blocks utilize the strengths of non-volatile devices for spin-based resistive/nanomagnetic storage to realize LUTs and SBs, as well as intrinsic computation by emerging devices such as summation, thresholding, etc. The CMOS logic blocks realize functional elements such as adders and multipliers. Whereas the inter-



Figure 3: HTCF hybrid fabric of spintronic and CMOS elements in HTCF Configurable Logic Block (H-CLB). H-CLBs comprise a fabric of Functional Blocks (FBs) & Switch Blocks (SBs).



Figure 4: Inter-device signal conversion transitions (left) and states (right): M=magnetization orientation, V=voltage, Q=charge, I=current.

device signal conversion requirements determined by the stateholding and state-changing mechanisms of these emerging/CMOS device blocks differ, signal conversion blocks are also encapsulated within the fabric. Thus, voltage-based switching devices such as transistors and current-based devices such as SHE or DWM devices undergo signal transformations using conversion-primitive circuit islands to allow fieldprogrammability. Interconnect points are integrated within a structured block to realize a flexible fabric of heterogeneous devices that retain their intrinsic signal representations. This allows workload-driven runtime composition of hardware resources to enable dynamic resiliency and energy tuning.

As depicted in Figure 4, inter-device signal conversion requirements can be determined by the state-holding and statechanging mechanisms of the devices under consideration. For example, the state diagram depicted in Figure 4 (left) identifies transitions to/from standard CMOS devices such as CLB input buffers, SRAM cells, and CMOS-based muxes that utilize a voltage-level representation. Meanwhile, the spintronic devices under evaluation utilize the magnetic orientation within a nanomagnet to represent the logic state. Various spintronic devices utilize distinct switching mechanisms involving current, voltage, or magnetic fields. The state-transition table in Figure 4 (right) depicts the state-to-signal conversions as three distinct translation methodologies: 1) Conversion of voltage *levels to magnetic orientation:* green arcs {a, c, e} correspond to signal-conversions whereby input voltages are applied to transistors that control current through the nanomagnet; 2) Generating voltage signals: corresponding to magnetic orientation: red arcs {b, d, f} indicate voltage signals generated based on orientation of a nanomagnet; 3) Translating external magnetic orientation to magnetic orientation-state: blue arcs $\{\mathbf{h}, \mathbf{g}, \mathbf{j}\}$ depict that the magnetization orientation of a nanomagnet can be correlated to an input magnetization. Conversion blocks are integrated with CLB designs consisting of LUTs, connection resources, and computational elements.

V. LOGIC PARADIGM HETEROGENEITY

An orthogonal dimension of fabric heterogeneity is also non-determinism enabled by either low-voltage CMOS or probabilistic emerging devices. It can be realized using probabilistic devices within a reconfigurable network to blend deterministic and probabilistic computational models. Herein, consider the probabilistic spin logic "p-bit" device [7] as a fabric element comprising a crossbar-structured weighted array. Programmability of the resistive network interconnecting p-bit devices can be achieved by modifying the resistive states of the array's weighted connections. Thus, the programmable weighted array forms a CLB-scale macro co-processing element with bitstream programmability. This allows field programmability for a wide range of classification problems and recognition tasks to allow fluid mappings of probabilistic and deterministic computing approaches. For example, a Deep Belief Network can be programmed in the field using recurrent layers of co-processing elements to form an $n \times m_1 \times m_2 \times ... \times m_i$ weighted array as a configurable hardware circuit with an *n*input layer followed by $i \ge 1$ hidden layers.

VI. CONCLUSION AND FUTURE WORK

The classes of reconfigurable architectures for emerging devices identified herein offer a feasible near-term platform and advantageous long-term pathway to pursue ITRS 2.0 themes 4 and 5. SCRFs embed a commercially-available replacement for SRAM cells, which can then be advanced towards HTCFs to realize a run-time configurable palette of heterogeneous technologies. Their diverse device technologies leverage the complementary features of: 1) volatility vs. non-volatility, 2) low switching energy vs. low leakage energy, and 3) soft-error susceptibility vs. radiation immunity, which are imparted by CMOS vs. emerging devices. These features can be exploited throughout a wide variety of applications ranging from low-cost / low-capacity / high-volume IoT adaptable platforms, up through high-capacity energy-aware FPGA-based accelerators for use in data centers. Current efforts involve developing transportable libraries, designing prototype LUTs and CLBs, and evaluating their energy, delay, and resilience for Boolean logic fabrics, stochastic spin-based probabilistic p-bit arrays, and neuromorphic reconfigurable fabrics [8].

REFERENCES

- T. M. Conte, E. P. DeBenedictis, P. A. Gargini, and E. Track, "Rebooting Computing: The Road Ahead," *IEEE Computer*, 50, 1, pp. 20-29, Jan. 2017.
- [2] S. Salehi, D. Fan, and R. F. DeMara, "Survey of STT-MRAM Cell Design Strategies: Taxonomy and Sense Amplifier Tradeoffs for Resiliency," *ACM Journal on Emerging Technologies in Computing*, Vol. 33, No. 3, pp. 1 – 16, April 2017.
- [3] W. Zhao et al., "New non-volatile logic based on spin-MTJ," *Physica Status Solidi (a)*, vol. 205, pp. 1373-1377, 2008.
- [4] P. E. Gaillardon et al., "Emerging memory technologies for reconfigurable routing in FPGA architecture," 17th IEEE Int'l Conf. Electronics, Circuits, and Systems, pp. 62-65, 2010.
- [5] D. Suzuki et al., "Area-efficient LUT circuit design based on asymmetry of MTJ's current switching for a nonvolatile FPGA," in *IEEE 55th International Midwest Symp. on Circuits and Systems*, pp. 334-337, 2012.
- [6] R. Zand, A. Roohi, D. Fan, and R. F. DeMara, "Energy-Efficient Nonvolatile Reconfigurable Logic using Spin Hall Effect-based Lookup Tables," *IEEE Transactions on Nanotechnology*, Vol. 16, No. 1, Jan. 2017.
- [7] K. Y. Camsari, R. Faria, B. M. Sutton, and S. Datta, "Stochastic p-bits for Invertible Logic," Physical Review X, vol. 7, p. 031014, 2017.
- [8] R. F. DeMara, M. Platzner, and M. Ottavi, "Guest Editorial: IEEE Transactions on Computers and IEEE Transactions on Emerging Topics in Computing Joint Special Section on Innovation in Reconfigurable Computing Fabrics from Devices to Architectures," *IEEE Transactions on Emerging Topics in Computing*, Vol. 5, No. 2, pp. 207-209, April-June 2017.