Heterogeneous Technology Configurable Fabrics: A Field-Programmable Paradigm for Leveraging Post-CMOS Devices in HPC

Ronald F. DeMara
Department of Electrical and Computer Engineering
University of Central Florida, Orlando, FL, USA 32816-2362, ronald.demara@ucf.edu

Abstract—Discussions are advocated regarding the merits of Heterogeneous Technology Configurable Fabrics (HTCFs) to realize the needed architectural abstraction as a pathway to manage Extreme Heterogeneity (EH) facing High Performance Computing (HPC) in the post-Moore era. HTCFs are reconfigurable logic arrays that leverage the cooperative roles of emerging spintronic and CMOS devices within a runtime adaptable platform residing as compute-node accelerators within post-exascale computing centers. HTCFs advance the challenges of EH by providing a uniform and familiar programming interface to multiple emerging device technologies. HTCFs leverage and extend the robust experience base and versatility of reconfigurable computing technology, theory, and tools to reduce barriers towards hardware implementation and mass adoption of post-CMOS devices in HPC. Like CMOS-based FPGAs, they assist minimizing communication occurrence and cost while reducing software and hardware bloat. Finally, it will be advocated that HTCFs also enable the new advance of “technology co-design” which trades off alternative mappings between emerging devices and CMOS technology during circuit synthesis-time for place-and-route optimizations, and also during runtime by remapping hardware resource locality to vital features such as non-volatility.

Keywords—Hardware Abstraction; Reconfigurable Computing Fabrics; Energy-Aware High Performance Computing; Dynamic Resource Optimization; Spintronics; Hardware Co-Design at Runtime.

I. NEW PERSPECTIVE TOWARDS EMBRACING HETEROGENEITY

It is advocated to debate the merits and implications of the vignette below as a transformative vision for post-Moore era HPC field-programmable accelerators. The perspective is to leverage both CMOS and post-CMOS devices for their relative strengths, such as fast switching speed, non-volatility, intrinsic addition/multiplication, and thresholding/neuromorphic computation which form a continuum of computational characteristics valuable to scientific computing. In particular, it is envisioned for Heterogeneous Technology Configurable Fabrics (HTCFs) to be available to the programmer and operating system as follows:

Dateline: Advanced Scientific Computing Research (ASCR) site, USA; March 1, 2040. A scientific programmer seeks to solve incompressible Navier-Stokes equations facing challenging resolution and accuracy requirements. She considers the processing phases involved and functional modules of the algorithms used along with several hardware attributes. These include a TFLOPS budget, data retention time / non-volatility, and a static vs. dynamic energy profile of the primary modules. She proceeds to map her design onto a palette of computational resources that include some selected spin-based devices within a high gate-equivalent capacity HTCF-component, as depicted in Figure 1. This HTCF component is functionally-identical to a Field-Programmable Gate Array (FPGA) based accelerator in use today, except it allows field programmable access and mapping to beyond-CMOS devices. This mapping occurs “in the field,” i.e. in the computational center, without fabrication of custom hardware in the foundry, similar field-programmable access to CMOS-based hardware.

She proceeds to enter source code which, as is already done today with FPGA-based translators via conversion to a Verilog functional description and configuration of IP blocks in a familiar FPGA-like specification, synthesis, and place-and-route toolchain interface. Technology-aware design optimizers then examine a set of functionally-equivalent, but technology-distinct design mappings realized as a field-programmable interconnection within a heterogeneous fabric. This overcomes two near-term hurdles: (1) CMOS-only designs face challenges

![Diagram](https://via.placeholder.com/150)

**Figure 1:** HTCF component vision. Physical devices consist of spintronic and CMOS elements within each HTCF Configurable Logic Block (H-CLB). H-CLBs comprise a larger field-programmable fabric of Functional Blocks (FBs) and Switch Blocks (SBs).
of technology scaling with regards to power and reliability, and (2) designers’ utilization of post-CMOS emerging devices requires extensive knowledge of emerging device physics, switching and interfacing behaviors, magnets, and an ASIC-only fabrication process. Fortunately, by having access to a reconfigurable HTCF component, she can validate her beyond-CMOS design immediately on the target hardware platform. Extensive non-recurring engineering design costs are also avoided. By encapsulating Extremely Heterogeneous (EH) device technologies within a reconfigurable fabric of compute accelerators, many advantages are worthy of discussion.

II. CHALLENGES & PERSPECTIVES OF POST-CMOS WITHIN HPC

Discussions of alternative pathways to support Beyond-CMOS devices are advocated to address the following Extreme Heterogeneity (EH) Challenges facing post-Moore HPC:

C1: attaining manageable and efficient resource utilization across disparate computing paradigms/components such as spintronic, CMOS, and stochastic computing-based accelerators having vastly different computation models and energy profiles,

C2: need for extensive modifications to existing FPGA accelerator toolchains in HPC to support post-CMOS devices.

Research Direction: Emerging post-CMOS devices comprise a vital aspect for future HPC accelerators. DeMara’s lab [1][2] and others at Stanford, Utah, EPFL France, AIMR Japan, Notre Dame, Lawrence Berkeley Lab [3], and elsewhere have been assessing the timing/energy/area of MRAM-based reconfigurable fabric elements via simulation with a focus on energy savings and reliability, while increasing future accessibility to emerging devices. DeMara has led such efforts and also a clearinghouse for simulation models, designs, and tools to be shared with the community worthy of discussion.

State of the Art: Significant results exist via simulation and fabricated prototypes for STT-MRAM based reconfigurable fabrics, as well as fabricated RRAM based fabrics, SPICE-based SHE-enhanced switching devices, and SOT-Neutral FPGA fabrics. These have included a number of benchmarking studies demonstrating results of 6% to 67% reduction in read energy, 21% reduction in reconfiguration energy, and 78% higher clock frequency versus alternative fabricated emerging reconfigurable architectures [2]. Other measurements of fabricated post-CMOS FPGAs have shown 81% power reduction over CMOS for representative applications, which can be achieved by leveraging the non-volatility feature of emerging resistive technologies [2]. A synopsis of recent innovations in reconfigurable computing ranging from devices to architectures has appeared in IEEE Trans. Comp. in 2017 [4].

Approach: Towards the realization of a feasible post-CMOS vision for HPC computing accelerators, initial talking points span features for programmers and tools to access the HTCF Configurable Logic Block (H-CLB) shown in Figure 1. H-CLBs form enabling elements to advance post-CMOS computing in HPC spanning Look Up Tables (LUTs), signal conversion circuitry, computation blocks, and fabric-scale interconnect, such as the three spin-based devices depicted. The potential benefits are summarized in Figure 1, which can be realized similarly to how field-programmablility enabled HPC advances during the CMOS design era. Another form of fabric heterogeneity is non-determinism via probabilistic devices [5] or clockless logic [6] by blending such computational models while abstracting device operational details from programmer.

III. POTENTIAL FOR HTCFs TO ADVANCE EH DEMANDS

Timeliness: Both challenges C1 and C2 are prudent to prepare for beyond-CMOS devices in HPC targeted for the 2025-2040 timeframe. MRAM devices as DRAM&SSD replacements are already commercially-available from Everspin/SK-Hynix, and recent presentations by ARM and AMD CTOs have elaborated on plans for MRAM-based devices in their products starting in the 2020-timeframe. Replacing SRAM FPGAs would be a next step in use of MRAM devices. Extensions to heterogeneous component arrangements were persistent trends in FPGAs over the past 20 years, and likely to continue in post-CMOS devices.

Uniqueness: HTCFs offer an intriguing vision for post-exascale architectures including novel device technologies which require tools and techniques for tailoring to the needs of scientists. As a fundamentally different way to compute, the mapping of operations to HTCF device technologies remains fluid. Flexible mappings become possible not only during circuit synthesis, but also during execution-time. Thus when scientific 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.

Novelty: Discussions of novel contributions to HPC result from crosscutting of device, circuit, and architectural levels spanned by next generation reconfigurable fabrics. These include: (i) new theories of non-charged based computing, (ii) new methodologies to encapsulate signal conversion between spin-based and charge-based Boolean representations, (iii) improved static versus dynamic energy profiles, and (iv) novel mechanisms for design time and runtime configurability of emerging devices. Knowledge will be generated to significantly advance post-CMOS design strategies to leverage larger-scale adoption of emerging devices. Fresh foundations spanning new computing paradigms, accelerator toolchain programming tools, and innovative emerging device uses for HPC acceleration beyond isolated fixed devices are in-scope of the discussion.

REFERENCES


