End-to-end Platform to Support the Democratization of System-scale Prototyping based on Emerging Devices

Gina Adam[1], Imtiaz Hossen[1], Osama Yousuf[1], Brian Hoskins[2], Advait Madhavan[2], Martin Lueker-Boden[3], Patrick Braganca[3], Vasileia Georgiou[3], Tiffany Santos[3]
ginaadam@gwu.edu
[1]*Electrical and Computer Engineering Department, George Washington University, Washington, DC 20052*
[2]*Alternative Computing Group, National Institute of Standards and Technology, Gaithersburg, MD, 20899, USA*
[3]*Memory & Networking Research Group, Western Digital Corporation, Milpitas CA, 95035, USA*

This work is focused on providing a robust prototyping framework that seamlessly integrates heterogeneous mixed-signal hardware and software environments for machine learning applications mapped to emerging device technologies. With this platform, novel two-terminal devices and algorithms can be easily benchmarked for speed, capacity and efficiency. The work demonstrates integration for RRAM and MTJ but the platform can easily be applicable to other emerging device technologies that have two terminals and fit the design specs on the CMOS chip.

New types of memristors and other emerging non-volatile memory devices based on novel designs and material stacks are constantly being researched and developed in academic facilities. While small-scale studies can be done with such hero devices, these academic efforts ultimately require in-house integration with separately fabricated foundry CMOS before being suitable for larger scale testing and prototyping[1]. Some device capabilities might only be apparent when integrated on CMOS[2]. However, building prototypes using emerging devices has many challenges in mixed signal data acquisition, hyperparameter optimization, and hardware co-processing. Foundry tape-outs are limited and expensive[3], and require a significant infrastructure for the circuit design, integration, and testing. Few researchers have the resources to develop their own.

This work proposes an end-to-end modular ecosystem for device benchmarking, capable of simulating and executing experiments on arrays of up to 20,000 two-terminal devices (Fig. 1). The system consists of complementary hardware and software environments[4]. The hardware environment (Fig. 2a) provided by a custom daughterboard supports the memristor/CMOS chip and interfaces directly to a field-programmable gate array (FPGA) development board. A software framework completes the system via a microprocessor hosted on the FPGA which supports an embedded Linux distribution and Python-based applications (Fig. 2b). Once an algorithm is validated in simulation and with real devices, individual methods could be replaced with FPGA hardware blocks or analog elements with the goal of eventually realizing the end-to-end training data-path in full ASIC implementation. This prototyping platform is planned to be open-sourced, with the long-term goal to facilitate a staged design process bringing together device researchers and computer engineers working on hardware for machine learning.

---

[1] F. Cai F et al. Nature Electronics **2**, 7, (2019), pp.290-299.
[2] Rao, M. et al. Nature, **615**, 7954, (2023), pp. 823-829.
[3] CMP France, NVMMAD200.
[4] B. Hoskins, et al. ICONS, 2021.

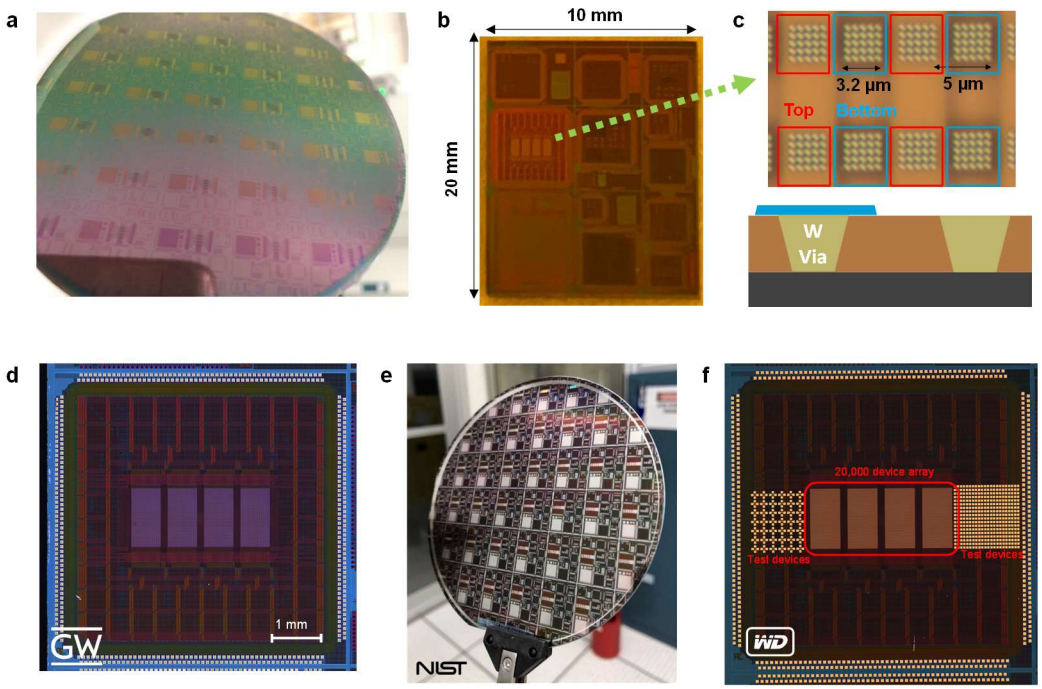

Figure 1: Emerging device integration on CMOS . (a) 4" CMOS wafer from foundry; (b) Chip diced out of the wafer with (c) open vias pre-planarized for device / CMOS integration; (d) memristor array integrated on chip via e-beam lithography at GW; (e) memristor array integrated on wafer via photolithography at NIST; (f) MTJ array integrated on wafer at WD.

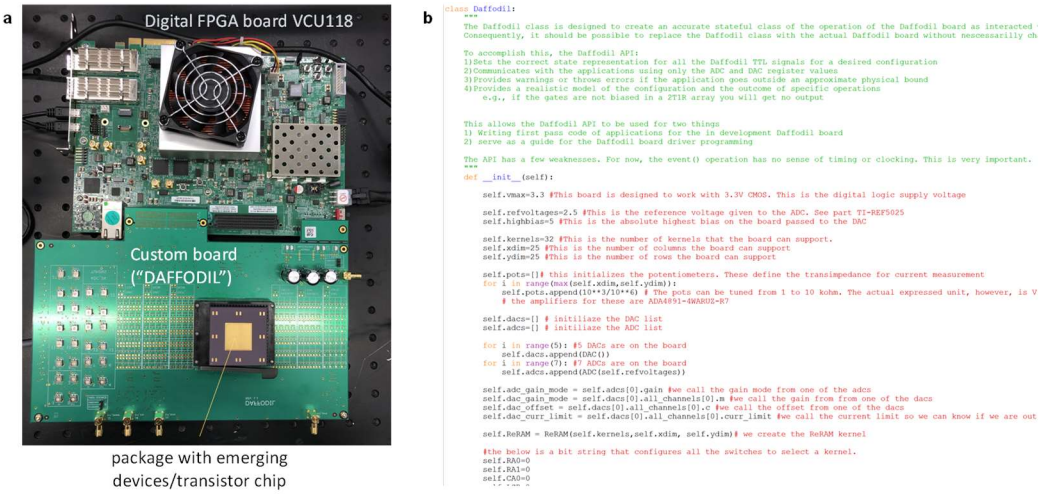

Figure 2: Prototyping platform. (a) End-to-end prototyping platform the packaged chip hosted by a custom board that interfaces directly with an FPGA board which can be connected to a computer for programming. (b) Software platform developed for the simulation and programming of the platform.