Progress Report (2023)

Analog-Mixed-Signal Computing Microcontrollers for Artificial Intelligence of Things: Toward Real-Time Carbon Monitoring and Management

1. Personnel:

- *Mingoo Seok, Associate Professor, Columbia University*
- *Yannis Tsividis, Professor, Columbia University*
- *Rentao Wan, current graduate student, Columbia University*

2. Project Period:

- Total period: Sep/1/2022 - Aug/30/2025 (4 years)
- Current: Sep/1/2022 – Aug/30/2013 (first year)

3. Summary of the Progress in the Current Period

The goal of the project is to create a new MCU integrating analog-mixed-signal (AMS) computing hardware for artificial intelligence of things (AIoT), which can show significant improvement in the energy efficiency and the throughput for artificial intelligence (AI) and machine learning (ML) workloads, such as convolutional neural networks (CNN).

In the past year, as a key building block, we have designed a computing-in-memory (CIM) macro with analog input, analog output, and analog bitcell. It performs multiply-and-accumulate (MAC) computation by discharging a capacitor C_{MAC} with the equation $\Delta V \cdot C_{MAC} = \sum[act_i T_u \cdot w_i I_u]$, where T_u and I_u are the unit time and the unit current, and act_i and w_i are activations and weights.

We explore two design techniques. First, we designed a custom 3-transistor 1-capacitor (3T1C) analog bitcell to store weights. Most previous CIM works use SRAM cells in macro, which need N cells to store N-bit weight, reducing the area and energy efficiency. The designed 3T1C analog cell stores a 4b weight as an analog voltage on a capacitor, achieving a smaller equivalent 4b weight area and higher weight storage density than other 28nm works. Also, to write an accurate weight-related voltage on the bitcell across process, voltage, and temperature (PVT) variation, we propose a current-based self-calibration writing scheme (CSCW). Previous work writes weights by providing a pre-determined constant voltage based on simulation results, which shows notable variability across PVT. During writing, CSCW builds a feedback loop with an OTA to tune the stored voltage until the discharging current harmonizes with the weight-corresponding writing current. In this way, each stored voltage fully accounts for variations, reducing the writing error greatly.

Second, this work combines various circuit techniques to maintain the written weight-related voltage, resisting capacitive-coupling noises and multiple leakages. We build the storage capacitor with a custom metal-oxide-metal (MoM) capacitor with a multi-finger structure on the top of bitcells to maximize capacitance density. Besides, we add an inverse signal path of the most serious coupling signal and an extra ground shield to restrict neighboring wires' coupling effects. Also, the transistors connected to the

storage capacitor in a bitcell are of minimum size to limit gate leakage. A tunable source bias is used to reduce the subthreshold leakage, and we keep the body of write-access transistors floating to limit the p-n junction reverse-biased leakage. All of these strategies significantly improve the retention time of the stored voltage.

We prototyped the test chip in 28nm CMOS technology. Compared with the state of the arts, it achieves the best energy efficiency and the best computes density for 4b MAC operation. Besides, the proposed macro shows a sufficiently low average normalized RMSE across variations.