

Case ID:M22-287P

Published: 7/15/2023

Inventors

Deliang Fan

Shaahin Angizi

Contact

Physical Sciences Team

Single-Cycle Processing-in-SRAM Logic Circuit Design

Traditional von-Neumann computing architectures, such as CPUs and GPUs, demonstrate limitations in memory bandwidth and energy efficiency. However, their high demand lies in their programmability and flexible functionality. Such platforms execute a wide spectrum of bit-wise logic and arithmetic operations. In this regard, recent application-specific processing-in-memory (PIM) designs suffer from the major challenge that their performance is intrinsically limited to one specific type of algorithm or application domain which means that such PIM platforms cannot keep pace with rapidly evolving software algorithms. To overcome this limitation, state of the art generic and programmable PIM architectures exploit alternatives to conventional bit-parallel algorithms. It is possible to realize arithmetic operations using bit-serial algorithms. However, it comes at a cost of high latency and more intermediate data write-back if multiple computing cycles are needed for basic in-memory Boolean logic functions. There is a need for a programmable processing-in-SRAM (PSRAM) accelerator that combines PIM computation efficacy with the programmability.

Researchers at Arizona State University have developed a programmable Processing-in-SRAM (PSRAM) accelerator chip design. The design is based on an 8T-SRAM array to accommodate a complete set of Boolean logic operations (e.g., NOR/NAND/XOR, both 2- and 3-input), majority, and full-adder, all in a single memory cycle. The design is implemented in a SRAM macro with size of 16 kb, demonstrating one of the fastest programmable in-memory computing system to date operating at 1.23 GHz. The 65 nm prototype chip achieves system-level peak throughput of 1.2 TOPs and energy efficiency of ~35 TOPS/W at 1.2 V.

Unlike prior multi-cycle processing-in-memory (PIM) logic designs that face multi-cycle operations, word-line underdrive, high-latency, read disturbance through modifying individual memory cells with large overhead or building look-up table-like operations by sacrificing memory function, this design has the following properties: (1) reconfigurable complete Boolean logic and majority function achievable in only one memory cycle; (2) full adder function achievable in only one ready cycle enabling more complex arithmetic computing such as multiplication; (3) highly parallel computing; and (4) no sacrifice of memory function and capacity.

Related publication: [A 1.23-GHz 16-kb Programmable and Generic Processing-in-SRAM Accelerator in 65nm](#)

Potential Applications:

- In-memory computing hardware platform

- Circuit design for in-memory computing
- PSRAM design for parallel vector operation, neural networks, data encryption, etc.

Benefits and Advantages:

- Provides the programmability required for in-memory computing platforms that could be used for various applications such as parallel vector operation, neural networks, and data encryption
- One-cycle in-memory Boolean logic design eliminates the redundant intermediate data write-back operations for 3-input logic and full adder that typically need multiple cycles with extra latency and energy
- Has been demonstrated on three applications: bulk bitwise vector operations, low-precision deep learning acceleration, and the Advanced Encryption Standard (AES) computation