Final Report

December 2025

Ian Barnaby
Ryn Stewart
Tome Dudanov

I. Problem Statement

In-memory computing to accelerate AI requires neural networks weights to be encoded in non-volatile memory (NVM). Current heterogeneous systems require weights to be communicated from CPU to accelerator, potentially in an untrusted environment. At present, there is no way to encrypt these weights in NVM to allow for correct neural network function. Securing this operation therefore (currently) requires encrypting/decrypting weights as communicated from CPU to accelerator to prevent snooping/monitoring attacks; however, as many of these applications are low-power, this proves difficult with a constrained area/power budget. This project seeks to describe a method of weight encryption in these systems leveraging existing device properties to reduce security overhead.

II. Motivation

Non-volatile memory (NVM) has emerged as a new memory paradigm in recent years, offering non-dynamic system memory that does not require refreshing. This type of memory has also been used for in-memory computing (IMC), as the two-terminal nature of certain NVM implementations (such as ReRAM [2]) allow for direct operation within the memory cells. This has become especially relevant for ML applications, in which neural networks may be directly mapped to a crossbar array of NVM cells [**?**]. Security in NVM cells is already a challenge due to data longevity, but has been researched using Bonsai-Merkle Trees (BMT) to secure the memory [**?**]. While effective in NVM for system memory, this is not an option for IMC, as the crossbar array must hold unencrypted values for correct neural network operation.

In the SLEAC project [1], the CPU/accelerator system consists of a CPU, IMC accelerator, and a data bus connecting the two. This system is prone to snooping attacks on the bus as weight values, along with their location within the array and instruction, are communicated from CPU to accelerator. Therefore, it is desired to determine a suitable encryption/decryption method for this unique low-power application such that instructions/data may be sent securely across the bus.

A. Threat Model

III. Experiment

IV. Evaluation

V. Related Works

References

[1] S. Davis, "SWAP Hub to Transform Satellite Imaging Performance Through Microelectronics-Enabled Artificial Intelligence - Semiconductor Digest — semiconductor-digest.com," https://www.semiconductor-digest.com/swap-hub-to-transform-satellite-imaging-performance-through-microelectronics-enabled-artificial-intelligence/, [Accessed 12-09-2025].

[2] H.-H. Hsu, T.-H. Wen, W.-S. Khwa, W.-H. Huang, Z.-E. Ke, Y.-H. Chin, H.-J. Wen, Y.-C. Chang, W.-T. Hsu, A. Lele, B. Zhang, P.-S. Wu, C.-C. Lo, R.-S. Liu, C.-C. Hsieh, K.-T. Tang, S.-H. Teng, C.-C. Chou, Y.-D. Chih, T.-Y. Chang, and M.-F. Chang, "A 22 nm floating-point reram compute-in-memory macro using residue-shared adc for ai edge device," *IEEE Journal of Solid-State Circuits*, vol. 60, no. 1, pp. 171 – 83, 2025/01/. [Online]. Available: http://dx.doi.org/10.1109/JSSC.2024.3470211