HNP-CGRA: A Hybrid Neuromorphic Processor with Coarse-Grained Reconfigurable Architecture Supporting Sparsity-Aware Acceleration

Yuehai Chen and Farhad Merchant

Bernoulli Institute and CogniGron, University of Groningen, The Netherlands

Spiking neural networks (SNNs) are a promising alternative to traditional artificial neural networks (ANNs) due to their binary spike-based computation and inherent sparsity. To exploit this sparsity efficiently, dedicated SNN accelerators have been widely studied in event-driven neuromorphic computing. However, such accelerators often lack flexibility and provide limited support for emerging SNN operators, especially in ASIC implementations [1]. In contrast, coarse-grained reconfigurable architectures (CGRAs) provide a favorable balance between flexibility and energy efficiency through word-level computation and runtime reconfigurability [2].

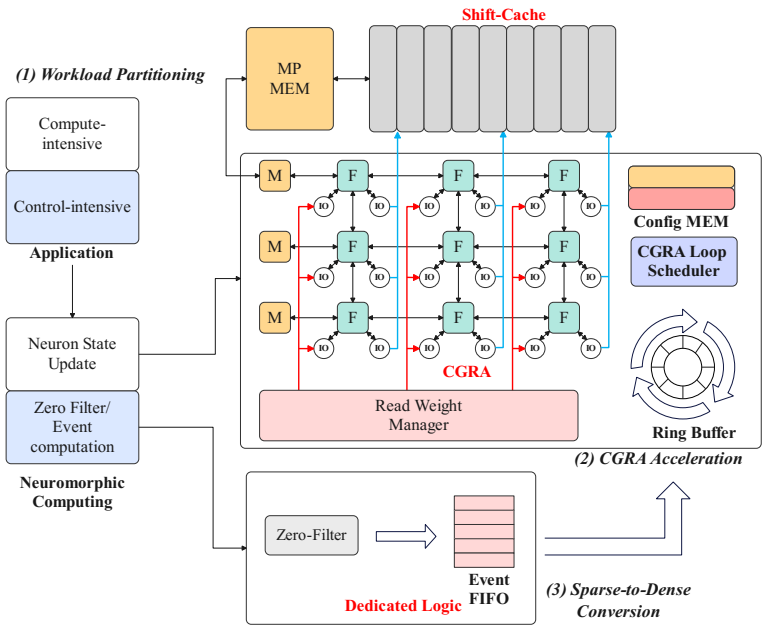

Figure 1: The overall architecture of HNP-CGRA

In this work, we propose HNP-CGRA, a hybrid neuromorphic processor tightly integrated with a CGRA for efficient SNN execution. As shown in Fig. 1, *(1) Workload Partitioning:* we first decompose neuromorphic workloads into compute-intensive and control-intensive components. *(2) CGRA Acceleration:* The compute-intensive part is mapped onto the CGRA, since different SNNs may involve diverse computational patterns; this not only improves computational efficiency for kernel execution, but also enhances scalability by flexibly supporting different operators. *(3) Sparse-to-Dense Conversion:* For the control-intensive part, which is not suitable for CGRA deployment, we design a dedicated event-filtering logic to perform sparse filtering and enable sparse-to-dense transformation, together with a corresponding control flow to coordinate the dedicated logic and the CGRA fabric. Finally, we deploy a deep spiking VGG-11 model on HNP-CGRA for the CIFAR-10 dataset. Experimental results show that HNP-CGRA achieves competitive performance while providing better flexibility.

[1] Yeh, Yao-Kai, et al. "A 16nm, 1Mb, 1-to-8b-Configurable 444.21 TOPS/W Fully Digital SRAM Compute-in-Memory Macro for Hybrid SNN-CNN Edge Computing." 2026 IEEE International Solid-State Circuits Conference (ISSCC). Vol. 69. IEEE, 2026.

[2] Ragheb, Omar, et al. "CGRA-ME 2.0: A research framework for next-generation CGRA architectures and CAD." 2024 IEEE International Parallel and Distributed Processing Symposium Workshops (IPDPSW). IEEE, 2024.