Dendritic based Computing Paradigm for Developing Scalable Neuromorphic AI Systems

PIs: Chiranjib Bhattacharya (CSA) & Chetan Singh Thakur (ESE)

Abstract: Artificial intelligence systems are increasingly constrained not by arithmetic throughput, but by data movement. In large-scale machine learning workloads, the dominant cost arises from transporting activations, gradients, and parameters across memory hierarchies and interconnect/network fabrics. As model sizes scale, communication latency, jitter, and congestion become first-order effects that degrade performance predictability and energy efficiency. In distributed neuromorphic and AI systems, the energy spent moving information often far exceeds the energy required for computation itself. For example, simulating one billion neurons on a large-scale system using TrueNorth processors is projected to consume approximately 4 kW of power, with only about 300 W attributed to computation and the remainder consumed by networking and communication infrastructure. At brain scale, the communication and configuration overhead rises to several hundred kilowatts, driving the effective energy utilization of such large-scale systems close to zero.

In contrast, the human brain operates at approximately 20 W of total power while supporting massively parallel, adaptive intelligence. Crucially, the brain computes asynchronously: neurons communicate via sparse, event-driven spikes and the biological interconnects —axons and dendrites —are not just passive wires. They perform elementary linear and nonlinear transformations such as low-pass filtering, coincidence detection, and amplification. In other words, communication pathways in the brain are computational substrates. This tight integration of computation and communication pushes the effective energy utilization of biological systems toward unity, offering a compelling blueprint for rethinking large-scale artificial intelligence architectures. These observations suggest that the traditional algorithmic abstraction—where computation and communication are decoupled—may no longer be optimal at scale.

This project proposes a radical shift in perspective: what if communication primitives themselves are treated as computational operators? The Processing-in-Interconnect (π^2) paradigm reformulates core machine-learning operations using primitives native to packet-switched systems—delay, time-based ordering, buffering, routing decisions, and event dropping. Instead of viewing these as transport mechanisms, we treat them as algorithmic building blocks. For example, weighted accumulation can be encoded through structured delay and temporal dynamics; nonlinearities can emerge through event dropping; and optimization dynamics can be expressed through adaptive routing and congestion feedback. The network is no longer a passive medium, but an active dynamical system whose behavior can be trained. The goal of this research is to develop a rigorous algorithmic framework for π^2 computing: formal models that map standard machine-learning objectives onto packet dynamics, differentiable abstractions that enable training, and software simulators that co-design algorithms with network behavior. By expressing AI workloads directly in terms of routing and timing primitives, π^2 aims to unify computation and communication at the algorithmic level. This perspective opens novel and unconventional directions in the field of neuromorphic computing and offers a more scalable architectural path toward achieving brain-scale AI training and inference.

Preferred background:

The ideal candidate should have a background in linear algebra, machine learning, and programming (e.g., Python). An interest in brain-inspired computing is essential.

References:

• K. D. Boahen, “Dendrocentric learning for synthetic intelligence,” Nature, vol. 612, no. 7938, pp. 43–50, Dec. 2022. doi: 10.1038/s41586-022-05340-6.
• M. Srivatsav, C. Bhattacharyya, S. Chakrabartty, and C. S. Thakur, “When routers, switches and interconnects compute: A processing-in-interconnect paradigm for scalable neuromorphic AI,” arXiv preprint arXiv:2508.19548, 2025.