A Family of FPGA-Validated Spiking Neural Network Processors with Hardware Cognitive Subsystems

Henry Shulayev Barnes
Catalyst Neuromorphic Ltd, Aberdeen, United Kingdom
henry@catalyst-neuromorphic.com

We present five generations of spiking neural network processors (N1–N5) designed, implemented in RTL, and validated on FPGA hardware. The processors span from an 8-core, 8,192-neuron baseline (N1) to a 512-core architecture with 4.2 million neurons, 8 neuron models, 10 synapse formats, and predictive coding (N4). All generations have been validated on Xilinx VU47P FPGAs via the AWS F2 platform. Measured FPGA performance: N1 at 62.5 MHz and 1.847 W, N2 at 62.5 MHz and 1.913 W (228,393 LUTs), N3 at 83.3 MHz and 1.923 W (262,317 LUTs). The N4-Edge 8-core variant achieves 15,668 timesteps per second with spike-to-output latency of approximately 69 microseconds.

N5 extends the architecture with 9 hardware cognitive subsystems implementing functionally distinct brain regions: basal ganglia (action selection via dopamine-modulated Q-learning), thalamic relay, cerebellar forward model, hippocampal spatial representation, prefrontal executive control, mirror neuron imitation, default mode network, amygdala (emotional valence), and metacognition. These are coordinated through a global workspace with ignition dynamics, coalition formation, and broadcast. A 5-state sleep-wake cycle (wake, N1, N2, N3, REM) gates sensory input and enables offline consolidation. N5 has been validated on FPGA with 273 tests passing across 5 independent runs, including demonstrated workspace ignition, Q-learning convergence, and sensory gating during sleep. To our knowledge, no other neuromorphic processor implements brain-region-level functional organisation in hardware.

The N4-Edge variant occupies 3,036 LUTs (2.6% of the Xilinx Kria K26) at 0.378 W total board power and 100 MHz with positive timing margin. ASIC projections at 28 nm place a 32-core N4-Edge at approximately 9.3 mm^2 and 100–300 mW.

On standardised temporal classification benchmarks using surrogate gradient training with int16 quantisation, the architecture achieves 91.0% on SHD (compared to Intel Loihi 2's 90.9%), 76.4% on SSC (Loihi 2: 69.8%), N-MNIST 99.2%, DVS Gesture 89.0%, and GSC-12 88.0%. Quantised inference retains 90.8% on SHD (0.2% degradation) and 76.4% on SSC (zero degradation).

The processors are accessible through a production Cloud API at api.catalyst-neuromorphic.com, providing commercially available access to neuromorphic hardware for the research community.

Keywords: neuromorphic computing, spiking neural networks, FPGA validation, cognitive architecture, global workspace, edge inference