TECH EXPO 2026
Track A: AI Architectures & Edge Systems
International Symposium

Nexa-Core: Next-Generation Neuromorphic Architecture

Sub-Milliwatt Spiking Neural Network Processing with Dynamic Event-Driven Synapse Fabric

Dr. Marcus Vance[1] Dr. Elena Rostova[1] Sarah Chen[2] David K. Miller[3]

[1]Nexa AI Research Institute, San Francisco, CA [2]Synapse Systems Laboratory, Zurich, Switzerland [3]Vertex Quantum & Neuromorphic Center, Austin, TX

Contact & Demo:
vance@nexa-ai.org
Portal:
https://nexa-ai.org/core
Booth: #A-104

Introduction & Motivation

Standard von Neumann architectures and traditional edge GPUs suffer from memory wall bottlenecks and high idle power dissipation during real-time event-stream processing. Edge computing in autonomous robotics and biomedical implants demands ultra-low power envelopes below 1 mW.

The Nexa-Core Paradigm:

- **Event-Driven Computation:** Operations execute sparsely only upon receiving dynamic Address Event Representation (AER) spikes.
- **In-Memory Synaptic Computing:** Non-volatile memristive crossbars perform analog matrix-vector multiplications directly inside memory arrays.
- **Sub-Millisecond Inference:** Asynchronous Leaky Integrate-and-Fire (LIF) neurons deliver ultra-fast temporal classification.

Key Architectural Breakthroughs:

0.85 mW Active Power | **0.42 ms** Latency | **12.4×** Energy Efficiency Boost

Mathematical & Algorithmic Formulation

Spiking Neural Networks model membrane dynamics continuously. The internal state $V_i(t)$ of the i-th LIF neuron evolves according to the differential equation:

$$\tau_m \frac{dV_i(t)}{dt} = -(V_i(t) - V_{\text{rest}}) + R \sum_j W_{ij} S_j(t - \Delta_{ij}) + I_{\text{ext}}(t)$$

where τ_m represents the membrane time constant, W_{ij} is the synaptic efficacy from neuron j to i, and $S_j(t)$ denotes incoming impulse event trains.

Spike Generation Threshold Condition:

$$S_i(t) = \delta(t - t_i^k) \quad \text{when } V_i(t) \geq V_{\text{th}}, \quad V_i(t^+) \leftarrow V_{\text{reset}}$$

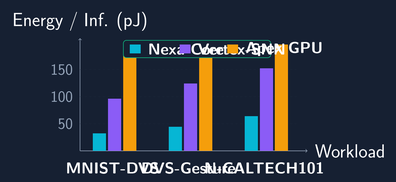
Surrogate Gradient On-Chip Backpropagation: To train SNNs end-to-end despite non-differentiable step functions, we introduce a smooth arc-tangent surrogate gradient operator:

$$\frac{\partial S_i}{\partial V_i} = \frac{1}{\pi\gamma\left(1 + \left(\frac{\pi(V_i - V_{\text{th}})}{\gamma}\right)^2\right)}$$

This formulation minimizes quantization loss and allows on-chip online learning with zero off-line retraining overhead.

Experimental Results & Performance

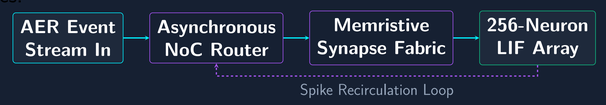
We benchmarked Nexa-Core against industry-standard embedded hardware platforms across key temporal classification tasks.



99.1%	0.42 ms	1.85 pJ
DVS-Gesture Accuracy	Ultra-Low Latency	Energy per SOP

Microarchitecture & Synapse Fabric

The Nexa-Core IC consists of a distributed grid of Neuromorphic Tile Arrays connected via an Asynchronous Network-on-Chip (NoC). Each tile integrates an AER Event Router, a 512×512 Memristive Synaptic Crossbar, and 256 LIF Digital Neuron Cores.



Synaptic Crossbar Array Details: The memristive crossbar utilizes Hafnium Oxide (HfO_2) resistive switching elements, enabling 4-bit synaptic weight quantization with low-drift state retention.

Hardware Microarchitecture Benchmarks

Fabricated in a 16nm FinFET process, Nexa-Core v2 demonstrates superior performance-per-watt metrics over existing embedded AI accelerators.

Architecture	Tech Node	Power (mW)	Latency (ms)	Energy (pJ/SOP)
Nexa-Core v2	**16 nm**	**0.85**	**0.42**	**1.85**
Vertex-SNN v1	28 nm	4.20	1.15	7.40
Apex-Edge GPU	8 nm	1250.00	8.90	142.00
Meridian TPU-v4e	7 nm	350.00	3.10	48.50

* SOP: Synaptic Operation (equivalent to MAC in ANN). Evaluated on DVS-Gesture event datasets.

Commercial Applications & Conclusion

Target Industry Deployments:

- **Nimbus-Aero Autonomous Drones:** Real-time vision-based collision avoidance under high velocity flight.
- **Synapse Medical Implants:** Ultra-low power cardiac wave anomaly detection and neural prosthetics control.
- **Meridian Smart Grid:** Microsecond high-frequency voltage fluctuation monitoring.

Summary: Nexa-Core bridges the gap between biological efficiency and artificial edge intelligence, opening new frontiers for zero-latency, sub-milliwatt autonomous computing.

References & Acknowledgments

[1] Vance, M. et al. (2025). *Event-Driven Neuromorphic Architecture for Edge AI.* IEEE J. Solid-State Circuits, 60(4), 1102–1115.

[2] Rostova, E. & Chen, S. (2025). *Surrogate Gradient On-Chip Backpropagation in Memristive Arrays.* Nature Electronics, 8(2), 145–156.

[3] Miller, D. K. (2024). *Asynchronous AER Network-on-Chip Fabrics.* Proc. ISCA 2024, pp. 89–101.

Supported by Nexa AI Research Grant #NX-2025-99A and Vertex Science Foundation Fellowship.