International Journal of Computational Research in Science and Technology
E-ISSN: 3108-0871, Volume 2 Issue 5
Sep-Oct-2026, PP 01-07

Neuromorphic Computing And Spiking Neural Networks For Energy-Efficient Edge Ai: A Review Of Models, Training Methods, Hardware And Applications

Kolhe Om Dinkar[1]; Prof. R. V. Varhade[2]; Prof. M. C. Kolse[3]; Prof. S. K. Said[4]
Department Of Artificial Intelligence And Data Science Jaihind College Of Engineering Pune, Maharashtra, India;

Abstract—Edge AI requires intelligent processing on resource-constrained devices such as wearables, drones, smart sensors and Internet of Things nodes. Conventional deep learning on von Neumann architectures consumes significant power because data must move continuously between memory and processing units, limiting its use in battery-powered applications. Neuromorphic computing addresses this challenge by emulating the structure of the biological brain through event-driven, massively parallel processing with memory co-located with computation. At its core are Spiking Neural Networks (SNNs), which communicate through sparse, discrete spikes over time so that energy is consumed mainly when meaningful events occur. This paper reviews the principles of neuromorphic computing and SNNs, covering neuron models such as the leaky integrate-and-fire neuron, spike encoding schemes, and training methods including ANN-to-SNN conversion, surrogate-gradient learning and biologically inspired local rules. Neuromorphic platforms such as IBM TrueNorth, the SpiNNaker system and Intel Loihi are compared, and edge applications in event-based vision, gesture recognition, keyword spotting and biomedical signal processing are examined. A layered architecture for neuromorphic edge AI is proposed, together with research gaps, objectives and a discussion of challenges such as training complexity, toolchain maturity and hardware standardization.

Index Terms—Neuromorphic computing, spiking neural networks, edge AI, energy efficiency, Loihi, TrueNorth, event-based vision, surrogate gradient

I. Introduction

Edge devices increasingly need to interpret sensor data on their own: a hearing aid that recognizes speech in noise, a drone that avoids obstacles, a smartwatch that detects an abnormal heartbeat. Running deep neural networks on such devices is constrained by battery capacity and heat. Horowitz [1] showed that fetching data from off-chip memory costs orders of magnitude more energy than an arithmetic operation, so the movement of weights and activations, rather than computation itself, dominates energy use in conventional processors.

The human brain performs perception and control on roughly twenty watts using about a hundred billion neurons that communicate with brief electrical pulses. Mead [2] proposed neuromorphic electronic systems that mimic the brain's organization, and Maass [3] described spiking neurons as a third generation of neural network models with greater computational power per unit than threshold or sigmoid units.

This paper reviews neuromorphic computing and SNNs for energy-efficient edge AI, covering models, training methods, hardware and applications, and proposes an architecture for neuromorphic edge systems.

II. Background and Fundamental Concepts

2.1 Spiking Neuron Models

The leaky integrate-and-fire (LIF) neuron integrates input current into a membrane potential that leaks over time; when the potential crosses a threshold, the neuron emits a spike and resets. In discrete time,

$$V[t+1] = \beta \cdot V[t] + \Sigma_i \, w_i \cdot s_i[t] - S[t] \cdot V_th \qquad (1)$$

where β is the leak factor, w_i the synaptic weights, $s_i[t]$ the incoming spikes and S[t] the output spike. Fig. 1 classifies the main elements of neuromorphic systems.

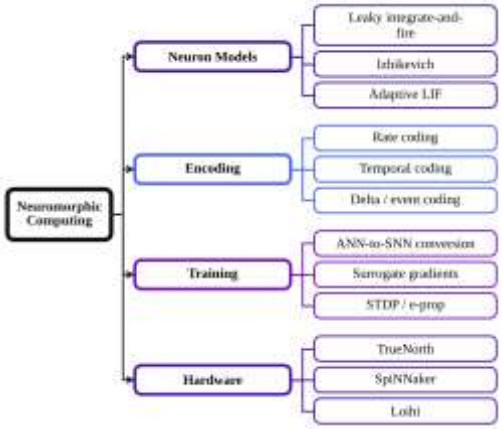


Fig. 1. Key elements of neuromorphic computing systems.

2.2 Spike Encoding

Real-valued data must be converted into spikes. Rate coding represents intensity by spike frequency; temporal coding uses the timing of the first spike; delta coding emits spikes only when a signal changes. Event cameras naturally produce delta-coded output, with each pixel reporting brightness changes independently [4].

III. Literature Survey

Hardware. Merolla *et al.* [5] presented IBM TrueNorth, a chip with one million spiking neurons and 256 million synapses organized in 4,096 cores, consuming tens of milliwatts during real-time operation. Furber *et al.* [6] described SpiNNaker, a massively parallel system of ARM cores connected by a packet-switched network optimized for spike communication, designed for large-scale brain simulation. Davies *et al.* [7] introduced Intel Loihi, a manycore neuromorphic processor with programmable on-chip learning, and later surveyed results showing large gains in energy and latency on suitable workloads [8].

Training. Because spikes are non-differentiable, standard backpropagation cannot be applied directly. Rueckauer *et al.* [9] converted trained conventional networks into SNNs by mapping activations to firing rates, and Sengupta *et al.* [10] extended conversion to deep VGG and residual architectures. Neftci *et al.* [11] explained surrogate-gradient learning, which replaces the derivative of the spike function with a smooth approximation, enabling direct training. Bellec *et al.* [12] proposed e-prop, a local learning rule suitable for on-chip learning in recurrent SNNs. Eshraghian *et al.* [13] provided a tutorial connecting deep learning practice with SNN training.

Applications. Amir *et al.* [14] built a fully event-based gesture recognition system using an event camera and TrueNorth, achieving low power and latency. Blouw *et al.* [15] benchmarked keyword spotting on Loihi and reported substantially lower energy per inference than on conventional processors. Esser *et al.* [16] showed that convolutional networks can run efficiently on neuromorphic hardware. Cramer *et al.* [17] introduced spiking audio datasets for systematic SNN evaluation. Roy *et al.* [18] and Schuman *et al.* [19] reviewed the field and identified algorithmic and application opportunities. Table I summarizes these works.

Ref.	Author(s) / Year	Technique / Approach	Key Contribution	Limitation / Gap
[5]	Merolla et al., 2014	TrueNorth million-neuron chip	Real-time inference at milliwatt power	No on-chip learning
[6]	Furber et al., 2014	SpiNNaker ARM-based system	Flexible large-scale spike simulation	Higher power than ASIC designs
[7]	Davies et al., 2018	Loihi manycore processor	Programmable on-chip learning	Research access; tooling maturity
[9]	Rueckauer et al., 2017	ANN-to-SNN conversion	Near-ANN accuracy on image tasks	Many time steps increase latency
[10]	Sengupta et al., 2019	Deep VGG and ResNet conversion	Scaled SNNs to deep architectures	Conversion-based limitations
[11]	Neftci et al., 2019	Surrogate-gradient learning	Direct gradient training of SNNs	Memory-heavy training over time
[12]	Bellec et al., 2020	e-prop local learning rule	Online learning for recurrent SNNs	Lower accuracy than BPTT
[14]	Amir et al., 2017	Event-based gesture recognition	Low-power, low-latency gesture system	Small gesture set
[15]	Blouw et al., 2019	Keyword spotting benchmark on Loihi	Lower energy per inference	Single benchmark task

TABLE I Summary of Literature Survey

IV. Research Gap

The following gaps are identified:

- SNN accuracy still trails conventional deep networks on many complex tasks.
- Training methods are computationally expensive and less mature than deep learning toolchains.
- Hardware platforms differ widely, and portable software frameworks are limited.
- Fair benchmarks comparing energy, latency and accuracy across platforms are scarce.
- On-chip continual learning for edge personalization is still at an early stage.

V. Objectives

The objectives of this review are:

1. 1) To explain the principles of neuromorphic computing and spiking neural networks.
2. 2) To review neuron models, spike encoding and SNN training methods.
3. 3) To compare major neuromorphic hardware platforms.
4. 4) To propose a layered architecture for neuromorphic edge AI systems.
5. 5) To examine applications, challenges and future research directions.

VI. Proposed System Architecture

The proposed architecture in Fig. 2 processes data as events end to end, from sensor to decision, so that energy is spent only when the input changes.

International Journal of Computational Research in Science and Technology
E-ISSN: 3108-0871, Volume 2 Issue 5
Sep-Oct-2026, PP 01-07

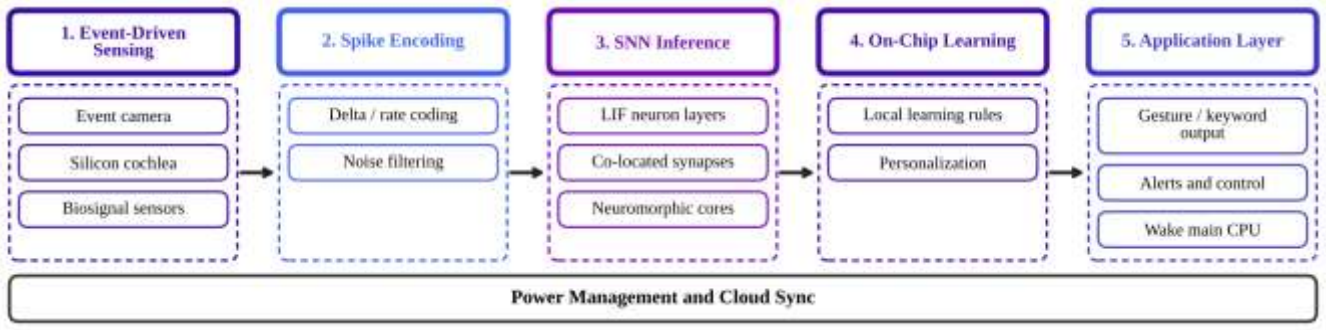


Fig. 2. Proposed layered architecture for neuromorphic edge AI.

6.1 Event-Driven Sensing Layer

Event cameras, silicon cochleas and spike-encoding front ends for conventional sensors produce sparse event streams.

6.2 Spike Encoding and Pre-processing Layer

Analog or digital signals are converted into spikes using delta or rate coding, and simple filtering removes noise events.

6.3 SNN Inference Layer

LIF neurons organized in convolutional or recurrent layers process spikes on neuromorphic cores. Synaptic weights are stored next to neurons, avoiding off-chip memory traffic.

6.4 On-Chip Learning Layer

Local learning rules adapt selected weights to the user or environment, for example personalizing a keyword spotter without sending data to the cloud.

6.5 Application Layer

Decisions trigger actions, such as waking a main processor, sending an alert or steering a drone. Results can be shared with the cloud when needed.

VII. Methodology and Working

The development workflow is shown in Fig. 3. Energy per inference on neuromorphic hardware can be approximated as

$$E \approx N_spikes \cdot E_syn + N_neurons \cdot E_upd \quad (2)$$

where N_spikes is the number of synaptic events, E_syn the energy per synaptic operation, N_neurons the number of neuron updates and E_upd the energy per update. Because sparse activity reduces N_spikes, energy falls as fewer events occur. In surrogate-gradient training, the non-differentiable step function is replaced in the backward pass by a smooth function such as

$$\partial S/\partial V \approx 1 / (1 + k \cdot |V - V_th|)^2 \quad (3)$$

International Journal of Computational Research in Science and Technology
E-ISSN: 3108-0871, Volume 2 Issue 5
Sep-Oct-2026, PP 01-07

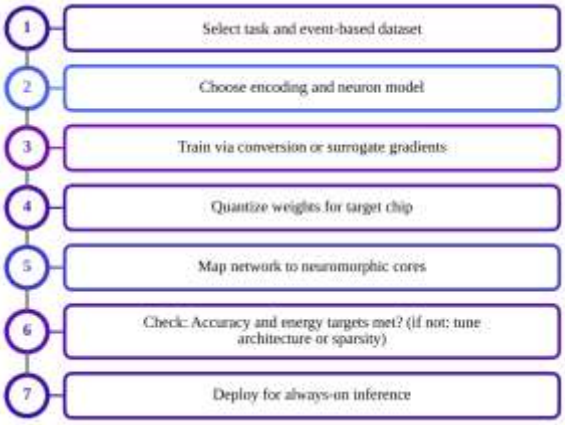


Fig. 3. Workflow for developing and deploying an SNN on edge hardware.

VIII. Comparative Analysis

Table II compares neuromorphic platforms. TrueNorth offers extremely low power but fixed functionality; SpiNNaker is flexible and suited to research; Loihi supports on-chip learning; and conventional GPUs remain best for training large models.

Platform	Learning	Power	Best Use
IBM TrueNorth	Off-chip	Very low	Fixed inference
SpiNNaker	Flexible (software)	Medium	Brain-scale simulation
Intel Loihi	On-chip	Very low	Adaptive edge inference
Edge GPU / NPU	Off-chip	Medium–High	Dense DNN inference
Microcontroller	Off-chip	Low	Tiny conventional models

TABLE II Comparison of Computing Platforms

IX. Challenges and Future Scope

Table III summarizes key challenges. Future directions include hybrid systems that pair conventional accelerators with neuromorphic co-processors, emerging memory devices such as memristors for analog synapses, standardized software frameworks, and large-scale benchmarks. Combining event cameras with SNNs is especially promising for drones and robots that must react within milliseconds on tiny power budgets.

Challenge	Mitigation
Training complexity	Surrogate gradients; conversion pipelines
Accuracy gap	Hybrid ANN–SNN; better neuron models
Toolchain maturity	Open frameworks (e.g., snnTorch, Lava)
Hardware diversity	Standard intermediate representations
Benchmarking	Shared energy–accuracy benchmarks

TABLE III Challenges and Mitigation

X. Conclusion

Neuromorphic computing and spiking neural networks offer a fundamentally different path to efficient AI by processing information as sparse events with memory and computation co-located. The reviewed literature shows that neuromorphic chips such as TrueNorth and Loihi achieve large energy and latency savings on suitable tasks, and that ANN-to-SNN

conversion and surrogate-gradient learning have narrowed the accuracy gap with conventional networks. For edge applications such as event-based vision, gesture recognition, keyword spotting and biomedical monitoring, neuromorphic approaches enable low-power, always-on intelligence, provided that training tools, hardware standards and benchmarks continue to mature.

10.1 Acknowledgment

The authors express their sincere gratitude to the Department of Artificial Intelligence and Data Science, Jaihind College of Engineering, Pune, for providing the facilities, guidance and encouragement needed to carry out this work. The authors also thank the faculty members and colleagues whose valuable suggestions helped improve the quality of this review.

References

1. M. Horowitz, “Computing’s energy problem (and what we can do about it),” in Proc. IEEE Int. Solid-State Circuits Conf. (ISSCC), 2014, pp. 10–14.
2. C. Mead, “Neuromorphic electronic systems,” Proc. IEEE, vol. 78, no. 10, pp. 1629–1636, 1990.
3. W. Maass, “Networks of spiking neurons: The third generation of neural network models,” Neural Netw., vol. 10, no. 9, pp. 1659–1671, 1997.
4. G. Gallego et al., “Event-based vision: A survey,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 44, no. 1, pp. 154–180, 2022.
5. P. A. Merolla et al., “A million spiking-neuron integrated circuit with a scalable communication network and interface,” Science, vol. 345, no. 6197, pp. 668–673, 2014.
6. S. B. Furber, F. Galluppi, S. Temple, and L. A. Plana, “The SpiNNaker project,” Proc. IEEE, vol. 102, no. 5, pp. 652–665, 2014.
7. M. Davies et al., “Loihi: A neuromorphic manycore processor with on-chip learning,” IEEE Micro, vol. 38, no. 1, pp. 82–99, 2018.
8. M. Davies et al., “Advancing neuromorphic computing with Loihi: A survey of results and outlook,” Proc. IEEE, vol. 109, no. 5, pp. 911–934, 2021.
9. B. Rueckauer et al., “Conversion of continuous-valued deep networks to efficient event-driven networks for image classification,” Front. Neurosci., vol. 11, Art. no. 682, 2017.
10. A. Sengupta, Y. Ye, R. Wang, C. Liu, and K. Roy, “Going deeper in spiking neural networks: VGG and residual architectures,” Front. Neurosci., vol. 13, Art. no. 95, 2019.
11. E. O. Neftci, H. Mostafa, and F. Zenke, “Surrogate gradient learning in spiking neural networks,” IEEE Signal Process. Mag., vol. 36, no. 6, pp. 51–63, 2019.
12. G. Bellec et al., “A solution to the learning dilemma for recurrent networks of spiking neurons,” Nat. Commun., vol. 11, Art. no. 3625, 2020.
13. J. K. Eshraghian et al., “Training spiking neural networks using lessons from deep learning,” Proc. IEEE, vol. 111, no. 9, pp. 1016–1054, 2023.
14. A. Amir et al., “A low power, fully event-based gesture recognition system,” in Proc. IEEE CVPR, 2017, pp. 7243–7252.

15. P. Blouw, X. Choo, E. Hunsberger, and C. Eliasmith, “Benchmarking keyword spotting efficiency on neuromorphic hardware,” in Proc. Neuro-Inspired Comput. Elements Workshop (NICE), 2019.

16. S. K. Esser et al., “Convolutional networks for fast, energy-efficient neuromorphic computing,” Proc. Natl. Acad. Sci. USA, vol. 113, no. 41, pp. 11441–11446, 2016.

17. B. Cramer, Y. Stradmann, J. Schemmel, and F. Zenke, “The Heidelberg spiking data sets for the systematic evaluation of spiking neural networks,” IEEE Trans. Neural Netw. Learn. Syst., vol. 33, no. 7, pp. 2744–2757, 2022.

18. K. Roy, A. Jaiswal, and P. Panda, “Towards spike-based machine intelligence with neuromorphic computing,” Nature, vol. 575, pp. 607–617, 2019.

19. C. D. Schuman et al., “Opportunities for neuromorphic computing algorithms and applications,” Nat. Comput. Sci., vol. 2, no. 1, pp. 10–19, 2022.