Neuromorphic Computing 2025: Current SotA

2025-09-01

Table of contents

1	Abs	stract	2
2	Intr	roduction	2
3	Har	rdware Advances (2019–2024)	3
	3.1	Digital Neuromorphic Chips	3
	3.2	Memristive Devices and Analog Neuromorphic Hardware	4
	3.3	Spintronic Neuromorphic Computing	4
	3.4	Photonic Neuromorphic Computing	5
	3.5	Neuromorphic Devices with 2D Materials	5
4	Alg	orithmic Advances (2019–2024)	6
	4.1	Spiking Neural Networks and Temporal Coding	6
	4.2	Surrogate Gradient Training for SNNs	7
	4.3	Biologically Plausible and Local Learning Rules	8
5	Opportunities and Gaps		9
	5.1	Toward Brain-Like Intelligence and AGI	9
	5.2	Neuromorphic Computing in Medicine and Healthcare	10
	5.3	Scientific and Industrial Applications	10
	5.4	Power-Efficiency and Scaling Challenges	11
	5.5	Integration with Conventional Computing	12
6	Cor	nclusion	13
7	Ref	erences	14

1 Abstract

Neuromorphic computing is a paradigm of designing hardware and algorithms inspired by the brain's architecture and principles, promising major gains in energy efficiency and new computing capabilities. This review provides a comprehensive overview of developments in neuromorphic computing from 2019 through 2024. We survey hardware advances - including digital neuromorphic chips (e.g. Intel Loihi, IBM TrueNorth, and SpiNNaker), emerging device technologies like memristors, spintronic circuits, photonic processors, and twodimensional (2D) material-based devices – that enable brain-like computation with vastly lower power than conventional electronics. We also summarize algorithmic advances in spiking neural networks (SNNs), covering progress in temporal coding strategies, the introduction of surrogate gradient methods for training SNNs like deep networks, and biologically plausible learning rules such as e-prop for online learning in spiking systems. Furthermore, we discuss key opportunities and gaps: the potential of neuromorphic systems to approach aspects of human cognition or artificial general intelligence (AGI), applications in medicine (like brain-machine interfaces and neural prosthetics) and science, the trade-offs between power efficiency and computational precision, and challenges in integrating neuromorphic accelerators into existing computing ecosystems. We conclude by highlighting how co-development of hardware and algorithms is critical to fulfill the promise of neuromorphic computing, and by outlining open research directions on the path toward more brain-like, efficient computing architectures.

2 Introduction

The human brain inspires a new class of computing architectures that radically depart from the conventional von Neumann paradigm. In a classical computer, memory and processing are separated, and operations occur in a sequential, clocked manner – a design that has led to tremendous performance gains but also faces power and scalability limits. Neuromorphic computing, first envisioned by Carver Mead in the 1980s, instead seeks to mimic the distributed, event-driven and parallel nature of brain networks. In a neuromorphic system, many simple processing units (artificial "neurons") operate in parallel and communicate via asynchronous spiking events, merging memory and computation locally at synapses. This design promises to circumvent the so-called von Neumann bottleneck (the limited bandwidth between processor and memory) by co-locating computation with memory, and to achieve dramatically higher energy efficiency akin to biological brains.

By 2019, neuromorphic computing had evolved from early analog circuits and small-scale prototypes into larger digital chips and emerging device technologies. Earlier milestones like IBM's **TrueNorth** chip (2014) demonstrated a fully digital neurosynaptic processor with 1 million spiking neurons, running on only about 70 mW of power. Likewise, the **SpiNNaker** system (first phase completed ~2018) incorporated a million ARM cores to simulate spiking neural networks in real time, primarily aimed at large-scale brain simulations. These efforts showed that orders-of-magnitude gains in energy efficiency are possible; TrueNorth, for example, delivered ~46 billion synaptic operations per second per watt in some tasks. However, they also underscored challenges: TrueNorth's neural model was rigid and difficult to program for complex tasks, and many academic neuromorphic platforms lacked software support, limiting their practical use.

Since 2019, research in neuromorphic computing has accelerated along two broad fronts. Hardware developments have diversified beyond digital CMOS chips into analog/mixed-signal designs and exotic technologies (memristive devices, spintronic circuits, photonic and 2D-material-based neuromorphic devices). These new hardware platforms aim to improve scalability, density, and bio-realism of neural computations. At the same time, algorithmic advances in spiking neural networks have made it easier to perform useful computations on neuromorphic substrates. Notably, researchers developed methods to train SNNs with high accuracy through surrogate gradients, and explored learning rules that are more biologically plausible or hardware-friendly (such as local synaptic plasticity and reward-modulated learning). Neuromorphic algorithms have also expanded into applications like vision, sensory processing, robotics, and optimization, often leveraging the event-driven nature of SNNs for real-time processing of streaming data.

This article reviews the key hardware and algorithmic innovations in neuromorphic computing over 2019–2024, and discusses the emerging opportunities and remaining challenges. We begin by surveying the state-

of-the-art neuromorphic hardware, from established digital chips to novel devices. We then cover progress in spiking neural network models, coding schemes, and learning algorithms that enable these systems to solve tasks. Finally, we examine how neuromorphic computing is being positioned in broader contexts – from efforts to approach brain-like intelligence, to use-cases in medicine and science – and identify gaps that must be addressed to realize the full potential of this paradigm.

3 Hardware Advances (2019–2024)

Neuromorphic hardware comes in many forms, but all share the goal of implementing neural network computations (weighted spikes, integrate-and-fire neurons, synaptic plasticity, etc.) directly in physics for superior efficiency. The 2019–2024 period saw significant progress in both **digital neuromorphic processors** and in emerging technologies that emulate neurons/synapses at the device level. Here we summarize advances in several major categories of neuromorphic hardware.

3.1 Digital Neuromorphic Chips

Digital CMOS neuromorphic chips use standard transistor technology to implement large spiking neural networks with user-programmable connectivity. These chips typically encode neuron states in digital logic but operate asynchronously and in parallel, often communicating via packet-based spike messages. **IBM's TrueNorth** was a landmark digital neuromorphic chip, with 4096 cores simulating 1 million spiking neurons and 256 million synapses on a single chip, while consuming mere milliwatts. TrueNorth proved that digital designs can achieve brain-like power efficiency; however, its neural model was fixed (e.g. no on-chip learning, limited precision) and running arbitrary networks on it was challenging.

Building on that foundation, researchers turned to more flexible digital architectures. The **SpiNNaker** project (University of Manchester and TU Dresden) developed a massively parallel computing platform with ARM cores that emulate spiking neurons in software. The second-generation **SpiNNaker-2** system, described in 2019, scaled up to a planned 10 million cores connected via a custom network. SpiNNaker 2 introduced adaptive power management features (dynamic voltage and frequency scaling, power gating) to allow energy use to scale with spiking activity. It also added hardware accelerators for tasks like convolution, making it useful not only for brain simulation but also for machine learning workloads. By exploiting a 22nm process and 3D integration, SpiNNaker-2 aims to increase spiking network simulation capacity by over 50× compared to its predecessor, opening doors to real-time simulation of networks with billions of synapses.

Meanwhile, Intel's Loihi neuromorphic chip (first released in 2018) matured into one of the most widely used research platforms in this period. Loihi featured 128 cores and around 130,000 neurons per chip, with fully digital yet highly flexible neuron models and on-chip spike-driven learning rules (e.g. spiking Hebbian updates). Critically, Intel provided a software toolchain for Loihi, enabling a community of over 100 research groups to experiment with it. By 2021, numerous results demonstrated Loihi's ability to solve tasks with significant speed and energy advantages over CPUs: e.g. constraint satisfaction problems, graph search, odor classification, and robotic control, often at orders of magnitude lower energy. These studies began to delineate niches where neuromorphic chips excel, such as sparse, event-driven workloads and problems requiring fine-grained parallelism. In late 2021, Loihi 2 was introduced with roughly 1 million neurons per chip and improved programmability (e.g. better support for dendritic compartments and higher precision), further advancing digital neuromorphic capabilities.

Overall, digital neuromorphic processors in 2019-2024 have shown that brain-inspired architecture can achieve extraordinary energy efficiency – often $100 \times$ to $1000 \times$ less energy per inference than conventional processors on suitable tasks. They also highlight a trade-off: digital chips offer speed and reliability, but reproducing rich neural dynamics or plasticity can be resource-intensive. This has led researchers to explore **mixed-signal and analog neuromorphic designs** and new device technologies that naturally emulate neuron/synapse behavior.

3.2 Memristive Devices and Analog Neuromorphic Hardware

One promising avenue for neuromorphic hardware is to use emerging memory devices (memristors, resistive RAM, phase-change memory, etc.) as artificial synapses and neurons. **Memristors** are two-terminal electronic devices that naturally remember their past current/voltage (via resistance state) and can thus implement synaptic weight storage co-located with computation. Over 2019–2024, significant progress has been made in integrating memristors into neuromorphic circuits. For example, researchers demonstrated large crossbar arrays of memristors performing analog matrix-vector multiplications in one step, effectively acting as layers of a neural network in hardware. Such in-memory computing leverages the physics of Ohm's law and Kirchoff's law: when input voltages are applied to rows, the currents summing at each column naturally compute the weighted sum through memristive conductances. This allows massively parallel, fast, and energy-efficient computation that bypasses the need to shuttle data between separate memory and CPU.

A 2024 review by Xiao et al. surveyed recent progress from fundamental memristive devices to full neuromorphic chips. Materials advances have produced memristors with high endurance (millions of cycles), retention, and multi-level analog states that are well-suited for representing synaptic weights. Novel devices such as phase-change memory (PCM) and spin-transfer torque magnetic RAM (STT-MRAM) have been used to build synaptic arrays that achieve online learning through local weight updates. For instance, researchers have demonstrated STDP (spike-timing dependent plasticity) and other local learning rules implemented in memristive crossbars, enabling unsupervised learning directly in hardware synapses. In parallel, prototype neuromorphic accelerators using memristor crossbars have been reported: e.g. hybrid CMOS-memristor chips that implement one or more layers of an SNN for tasks like image recognition. These systems exploit analog computing for the core dot-products, with digital logic handling peripheral functions (thresholding, resets, communication).

By bringing memory and computation together, memristive neuromorphic hardware can attain tremendous energy efficiency and density. However, a key challenge highlighted in recent studies is device variability and imperfections. Analog computing with nanoscale devices inevitably introduces noise and variability in weights, which can degrade accuracy. Techniques like differential encoding, calibration, or training algorithms robust to analog noise have been developed to mitigate these issues. Despite challenges, the consensus is that memristor-based neuromorphic hardware holds great promise for fast, low-power AI at the edge, as evidenced by progress in the early 2020s. By 2024, memristive neuromorphic prototypes were tackling tasks in vision, speech, and associative memory with competitive accuracy, pointing toward integration of these devices in future neuromorphic co-processors.

3.3 Spintronic Neuromorphic Computing

Spintronic neuromorphic devices leverage the spin of electrons and nanomagnetic phenomena to mimic neural behavior. Spintronic devices are inherently non-volatile (retaining state without power) and can exhibit dynamics such as oscillations and threshold switching that parallel neuron spiking. A 2024 review by Marrows et al. surveys the state-of-the-art in using spintronic technology for neuromorphic computing. Key components include spintronic synapses – often based on magnetic tunnel junctions (MTJs) whose resistance can be tuned analogously to a synaptic weight – and spintronic neurons, such as voltage-controlled or current-controlled oscillators that can produce spiking-like outputs.

Recent work has shown that MTJ-based memory cells can serve as efficient synapses: for example, one group built an associative memory chip with MTJ-based logic-in-memory that achieved ~90% power reduction compared to a conventional CMOS design. Meanwhile, prototypes of spintronic neurons have been implemented using devices like spin-torque nano-oscillators, which can integrate input currents and exhibit non-linear dynamics analogous to integration-and-fire. These devices naturally operate in the analog domain and can oscillate in the GHz range, potentially enabling very fast neural computing. A popular approach in spintronics is **reservoir computing**, where networks of coupled oscillators or domain-wall nanowires process information without needing precise weight tuning. Several demonstrations in 2019–2023 used spintronic reservoirs to perform tasks like pattern recognition and time-series prediction with competitive accuracy, albeit often requiring external post-processing.

The scientific outlook is that spintronics could offer extremely high-speed and low-power neuromorphic devices, with features like inherent stochasticity (useful for probabilistic computing) and easy integration with existing CMOS for hybrid systems. However, many spintronic neuromorphic components are still in early research stages: achieving reliable, large-scale integration of millions of spin devices with controllable behavior is a work in progress. The 2024 Marrows et al. review concludes that significant advances in materials (to improve device uniformity and reduce noise) and circuit architectures will be needed to bring spintronic neuromorphic computing to practical utility. If those advances occur, spin-based neuromorphic hardware could complement or even surpass CMOS in specific applications due to its non-volatility, rich dynamics, and potential for analog processing at very low energy cost.

3.4 Photonic Neuromorphic Computing

Optical or photonic neuromorphic computing emerged in this period as an exciting approach to achieve ultrafast neural networks by using light rather than electrical signals. Photonic systems can leverage the high bandwidth of optical signals and the natural parallelism of light propagation to perform neural computations with potentially sub-nanosecond latencies. A comprehensive 2024 review by Li et al. in Advanced Materials highlights the growth in **integrated photonic neuromorphic systems**. In these systems, components like microring resonators, Mach-Zehnder interferometers, phase-change materials, and semiconductor lasers act as neurons and synapses on photonic chips. For example, an optical neuron can be implemented by a laser that emits a pulse when its input optical intensity exceeds a threshold, analogous to spiking, while synaptic weights can be tuned via optical modulators or material phase states that affect light transmission.

Between 2019 and 2024, researchers demonstrated photonic circuits that implement small neural networks for tasks like image recognition, logic operations, and signal processing. One notable result was a photonic convolutional accelerator that used wavelength-division multiplexing to perform many dot-product operations in parallel across different colors of light. By slicing broadband light into multiple channels, the system achieved high parallelism and performed convolutional neural network inference optically at speeds beyond GHz rates. Another development was the use of **phase-change photonic memory** devices: by using materials like GST (commonly used in optical storage) integrated on waveguides, weights of a photonic neural network could be stored and applied directly in the optical domain with low energy.

Photonic neuromorphic computing offers key advantages: very high speed (because light travels fast and operations like interference occur essentially at light-speed) and no ohmic losses, which suggests lower energy per operation at large scale. However, current photonic neural networks are limited in size and precision. The devices can be bulky (on-chip photonics still require micrometer-scale components) and controlling them with precision is difficult due to sensitivity to temperature and fabrication variation. There is active research into more compact photonic devices (e.g. using nanophotonics, metasurfaces) and better integration of photonics with electronics for control. The 2024 Li et al. review concludes that while photonic neuromorphic hardware has made great strides – moving from individual photonic neurons to integrated neural network prototypes – several "breakthrough" device innovations and co-design of optical systems with neural algorithms are needed to realize its full potential. Still, the prospect of optical processors executing neural network inference or learning tasks at terahertz bandwidths with minimal heat dissipation remains a compelling long-term goal.

3.5 Neuromorphic Devices with 2D Materials

Two-dimensional materials (atomically thin semiconductors like graphene, MoS, h-BN, etc.) have attracted attention for neuromorphic computing due to their unique electrical properties and suitability for dense integration. Researchers have explored **2D material-based memristors and transistors** to function as artificial synapses and neurons, often termed "memtransistors" when a single device can exhibit combined memory and transistor behavior. A 2025 review by Choi *et al.* provides a comprehensive overview of neuromorphic systems based on 2D materials.

One advantage of 2D materials is their atomic thickness, allowing extreme scaling and even *vertical* integration of multiple layers. For instance, van der Waals heterostructures can stack different 2D materials to

create synaptic devices with multiple programmable conductance states and built-in memory functionality. These can serve as multi-bit synapses or dynamic synapses that emulate short-term and long-term plasticity phenomena. Additionally, 2D semiconductors like MoS offer high carrier mobility and subthreshold device operation, enabling transistors that switch with very low voltage – a boon for energy-efficient neuronal circuits. The flexibility of 2D materials also allows for potential neuromorphic sensors and processors on flexible substrates, opening the door to wearable or implantable neuromorphic chips.

During 2019–2024, multiple proof-of-concept neuromorphic devices using 2D materials were demonstrated. Examples include MoS -based memristors showing STDP behavior (where the conductance change depended on the relative timing of voltage spikes applied, akin to biological synapses) and Black Phosphorus or WSe transistors that could integrate pulses and fire, mimicking neuron spiking. Some works achieved all-2D neuromorphic circuits – e.g. an array of graphene synapses coupled with MoS neuron transistors that together performed pattern recognition with online learning. The 2025 review by Choi et al. also emphasizes the potential for monolithic 3D integration using 2D materials: because these materials can be layered without destroying each other's properties, one can envision stacking tens of layers of neurons and synapses in a single chip, dramatically increasing density beyond what 3D transistor stacking allows. Such vertical neuromorphic circuits could emulate the brain's dense interconnectivity in a compact footprint, something traditional silicon struggles with due to heat and fabrication constraints.

The field of 2D neuromorphic devices is still nascent, but the unique characteristics of 2D materials – atomic scale thickness, surface-driven properties, flexibility – offer complementary advantages to conventional technology. Challenges remain in achieving uniform, reproducible devices and integrating them into large-scale circuits (many demonstrations are of single devices or small arrays under carefully controlled conditions). Nonetheless, by 2024 researchers have identified clear opportunities where 2D materials could push neuromorphic hardware forward, especially in scenarios requiring extreme density (e.g. 3D integrated crossbar networks) or interfacing with biology (flexible, biocompatible neural interfaces). As fabrication and material synthesis methods improve, 2D-material neuromorphic systems may become a key piece of the broader neuromorphic computing landscape.

4 Algorithmic Advances (2019–2024)

While hardware provides the platform, algorithms and models determine what neuromorphic systems can do. In the past few years, there has been considerable progress in spiking neural network algorithms, inspired by both neuroscience and deep learning. These advances seek to make SNNs more trainable, more efficient, and more capable of performing complex tasks. Here we highlight three important areas of algorithmic development: improvements in spiking neural network models and coding schemes, the advent of surrogate gradient techniques for training SNNs, and progress in learning rules that bring SNN training closer to biological plausibility without sacrificing performance.

4.1 Spiking Neural Networks and Temporal Coding

Spiking neural networks (SNNs) are the primary model of computation in neuromorphic systems. Unlike traditional artificial neural networks that use continuous-valued activations, SNNs communicate via discrete spikes (events) over time. Each neuron integrates incoming spikes and emits its own spike when its membrane potential exceeds a threshold, potentially after some delay. The **temporal dimension** of spikes – i.e. not just how many spikes are fired, but when they fire – can carry information. An ongoing research question has been how to best encode information in SNNs: through **rate coding** (where the spike rate over an interval corresponds to a value) versus **temporal coding** (where precise spike timings or spike order convey information). Temporal coding schemes exploit the high timing precision possible in neuromorphic hardware and potentially enable faster, more efficient computation than rate-based approaches.

Prior to 2019, most SNN applications adopted simple rate coding to leverage existing deep learning methods (by converting pre-trained analog neural networks into spiking ones). However, 2019–2024 saw a surge of

interest in temporal coding for SNNs, because of its potential to make use of each spike's information content more effectively. For example, researchers like Mostafa showed that if one uses **time-to-first-spike** as the code (where a neuron fires sooner for a larger input), the input-output mapping of a spiking network can become differentiable and thus trainable with gradient descent. In a 2018 study, Mostafa demonstrated supervised learning in a feedforward SNN using temporal coding, achieving high accuracy on MNIST with far fewer spikes than rate-based networks, since each neuron just fired at most one spike per input example. This highlights a key advantage: SNNs with temporal codes can be extremely sparse in their spiking (many neurons remaining silent unless needed), which translates to energy efficiency on neuromorphic hardware.

Other works have explored rank-order coding (where the rank of a neuron's spike time among peers encodes value) and phase coding (where spikes fired relative to a global oscillation phase carry information). Temporal coding is also naturally utilized in **event-based sensory processing** – for instance, neuromorphic vision sensors (event cameras) output spikes when pixels change, so SNNs processing these data inherently operate on the timing of incoming events. By the early 2020s, SNNs were achieving impressive results on tasks like event-based vision (e.g. classifying hand gestures or driving scenes from event camera input) by leveraging spatiotemporal patterns of spikes rather than averaging them into rates.

However, temporal coding introduces challenges for learning, since the exact spike timing is a non-differentiable and discontinuous variable. This motivated new training methods (discussed next) to handle such cases. It's worth noting that the field has not settled on a single "best" coding scheme – instead, the coding may be task-dependent. What has become clear is that neuromorphic hardware and SNNs excel in scenarios where information is naturally event-driven or time-dependent (audio streams, sensor signals, etc.), and using the temporal structure of spikes (rather than forcing them into average rates) can unlock better efficiency and low-latency processing that would be hard to replicate in traditional networks. This aligns well with the real-world, where stimuli often arrive as asynchronous events; SNNs can process and respond to each event in real-time, rather than accumulating data into frames or batches.

4.2 Surrogate Gradient Training for SNNs

A major breakthrough in the late 2010s that carried through the early 2020s was the development of surrogate gradient methods for training spiking neural networks. The core difficulty in training SNNs with backpropagation is that neuron spike events are not differentiable; the Heaviside step function used to decide spiking has zero derivative almost everywhere and infinite derivative at threshold, which breaks gradient-based optimization. Earlier approaches to train SNNs either avoided true spikes (using rate-based approximations) or relied on biologically inspired local rules (like STDP), which did not reach the accuracy levels of deep learning on complex tasks.

Surrogate gradient learning addresses this by replacing the non-differentiable spike function with a smooth surrogate function during the backward pass. Essentially, one defines an approximate gradient for the spike – for example, using a fast sigmoid or triangular function as a stand-in for the spike's step – which allows gradients to flow through the network during training. This trick enables the use of backpropagation-throughtime (BPTT) on SNNs, treating them similarly to recurrent neural networks. Emre Neftci and colleagues demonstrated that with surrogate gradients, SNNs could be trained on image classification tasks to nearly match the performance of non-spiking networks, all while maintaining sparse spiking activity. This result was pivotal: it brought SNNs into the realm of high-accuracy deep learning, rather than being limited to toy problems or requiring manual tuning.

Following this, many works refined surrogate gradient techniques. Surrogate functions were crafted to balance accuracy and biological plausibility – for instance, one might choose a surrogate that is non-zero only near the threshold, mimicking the idea that only near-threshold events affect learning, which has some neurobiological grounding. By 2022, surrogate gradient-trained SNNs achieved state-of-the-art or near state-of-the-art on benchmarks like CIFAR-10 and even ImageNet (using deep spiking convolutional architectures), sometimes using fewer time steps or spikes than earlier attempts. A notable example is the work of Fang and colleagues who built deep residual SNNs and applied surrogate gradient training with additional techniques (normalization, data augmentation) to reach high accuracy on CIFAR-10 with very low latency (5–10 simulation steps) – something that wasn't conceivable a few years prior.

The impact of surrogate gradients is that it unlocked gradient-based **end-to-end training** for SNNs, much like backprop did for traditional neural nets decades ago. This means one can optimize spiking networks for arbitrary loss functions and tasks, making neuromorphic hardware much more programmable and application-versatile. The trade-off is that backprop-through-time on SNNs, especially with long simulation durations, can be memory and computation heavy on conventional hardware (since it unfolds the temporal dynamics). Researchers are actively exploring more efficient training, including *batchless* online training suited for neuromorphic hardware deployment. Nonetheless, surrogate gradient learning stands as a cornerstone advance, bringing together the efficiency of spike-based computation with the powerful training algorithms of deep learning. As a result, the algorithmic gap between SNNs and ANNs has significantly narrowed from 2019 to 2024, making it feasible to tackle complex pattern recognition or motor control tasks with SNNs and achieve comparable accuracy, while potentially reaping energy advantages when such networks run on neuromorphic chips.

4.3 Biologically Plausible and Local Learning Rules

While surrogate gradients borrow from machine learning, another thread of research pushes for training methods that are more aligned with biological mechanisms or more amenable to on-chip learning. One highlight in this area is the concept of eligibility propagation (e-prop) introduced by Bellec et al. in 2020. E-prop is an algorithm for training recurrent spiking networks using only information locally available at each synapse and node (unlike backpropagation which requires global information). In their Nature Communications paper, Bellec and colleagues showed that e-prop can approach the performance of backprop-through-time on tasks like speech and music recognition, without requiring the full sequential unfolding of BPTT. E-prop works by computing eligibility traces at each synapse as the network runs – these traces capture a synapse's recent contributions to network activity. A global feedback signal (like a reward or error broadcast, which could be dopamine-like in the brain) then modulates these traces to perform weight updates. Crucially, the heavy lifting in credit assignment is done locally and online, making it far more plausible as a model of learning in biological circuits and more amenable to implementation on neuromorphic hardware that can support local plasticity.

In the 2019–2024 period, e-prop and related approaches (such as various forms of spike-based reinforcement learning, or approximations to backprop using locality constraints) have gained traction. For example, one study integrated e-prop on the SpiNNaker-2 platform and demonstrated on-chip learning for a spiking recurrent network, showcasing that even without a traditional compute cluster, neuromorphic hardware could *learn* from data in real-time using local rules. Other researchers extended these ideas by combining local learning rules with **neuromodulatory signals** – akin to brain's reward systems – to enable one-shot or few-shot learning in SNNs for tasks like navigation and adaptation.

Another biologically-inspired learning mechanism is spike-timing-dependent plasticity (STDP) and its variants. STDP is an unsupervised rule adjusting synapses based on the relative timing of pre- and post-synaptic spikes (strengthening connections when a pre-synaptic spike precedes a post-synaptic spike by a short interval, and weakening otherwise). While STDP alone is often not sufficient for complex tasks, variations of it have been used in combination with reinforcement signals or used to self-organize network feature detectors. In the early 2020s, researchers developed **hybrid learning** approaches: for instance, using STDP to pre-train layers of an SNN (to learn useful feature representations in an unsupervised way), then fine-tuning the network with surrogate gradient supervised training. Such hybrid approaches can leverage the best of both worlds: efficient unsupervised adaptation and high-accuracy task-specific learning.

Overall, the drive toward more biologically plausible learning is motivated by both scientific curiosity (understanding how real neural circuits might learn) and practical considerations (enabling online learning and lifelong adaptation on neuromorphic devices without needing cloud computing). The progress in algorithms like e-prop suggests that one can achieve near-backprop performance with local learning rules. This is encouraging for future neuromorphic systems that might continuously learn from their environment (e.g. a robotic agent adapting on the fly) – something currently infeasible with large deep learning models that require offline retraining. It's worth noting that there is still a gap between what is plausible in a biological sense and what is maximally efficient in an engineering sense; thus, research continues into algorithms that

strike different balances along that spectrum. The period up to 2024 has provided a rich toolbox of learning methods for SNNs, ranging from pure engineering-driven (surrogate gradients) to bio-inspired (STDP, e-prop, Hebbian learning), and demonstrated that each has its domain of applicability.

5 Opportunities and Gaps

Neuromorphic computing sits at the intersection of computer engineering, neuroscience, and machine learning. As we survey the achievements of 2019–2024, it becomes evident that neuromorphic systems have advanced substantially, yet they have not entirely revolutionized computing... at least, not **yet**. In this section, we discuss the key opportunities that lie ahead for neuromorphic computing – the domains where it could be uniquely transformative – and the gaps or challenges that must be addressed to realize these opportunities. Topics include the quest for brain-like intelligence, applications in medicine and science, the perennial issue of power efficiency versus performance, and the integration of neuromorphic hardware into mainstream computing infrastructure.

5.1 Toward Brain-Like Intelligence and AGI

A long-term aspiration (and oft-used justification) for neuromorphic computing is to move us closer to artificial general intelligence (AGI) by adopting the brain's computing principles. Spiking neural nets with plastic synapses are arguably closer to biological neural networks than the static, dense layers of deep learning. Could neuromorphic systems one day exhibit cognitive abilities rivaling biological brains? This remains an open question and a driving motivation. Roy et al. (2019) noted that while neuromorphic hardware achieves impressive efficiency, a major open problem is task generalization – applying learned knowledge to new situations – something at which the brain excels but current AI struggles. Neuromorphic architectures alone do not guarantee general intelligence, but they do enable experimentation with large-scale models of the brain (e.g. models with spiking neurons, dendritic compartments, and local learning) that might shed light on principles of intelligence.

One opportunity is in large-scale brain simulation and brain-inspired algorithms. Projects like the European Human Brain Project have leveraged neuromorphic platforms (SpiNNaker, BrainScaleS) to simulate cortical microcircuits in hopes of understanding neural computation. Although early results fell short of major discoveries, continuing improvements in hardware and models may eventually allow simulation of neural systems at unprecedented scales and realism. On the algorithmic front, neuromorphic systems naturally support forms of computation that deep networks find difficult – for example, spike-based probabilistic sampling, dynamic adaptation and learning, and sparse, event-driven sensing. These properties could be important pieces in the AGI puzzle.

However, a gap remains between current neuromorphic capabilities and the requirements of open-ended general intelligence. Today's neuromorphic chips and SNN models, while brain-inspired, are still far simpler than the brain in structure and function. Cognitive functions like reasoning, language, and abstraction have not been demonstrated on SNNs at anywhere near the levels achieved by large deep learning models. There is an ongoing debate whether closing this gap requires more sophisticated neuromorphic designs (incorporating features like complex dendrites, neuromodulators, etc. found in biology) or if current hardware could do more if paired with the right learning algorithms. In either case, neuromorphic computing provides a complementary path to mainstream AI: focusing on architectural efficiency and online adaptation rather than sheer scale of data and parameters. The opportunity is that by exploring this path, we might discover new computational paradigms that contribute to AGI, or at least to more general and adaptive AI systems. The next decade will likely see increasing synergy between neuromorphic engineering and fields like cognitive science and robotics, as researchers attempt to imbue these systems with higher-level functionality.

5.2 Neuromorphic Computing in Medicine and Healthcare

One area where neuromorphic technology shows significant promise is **biomedical applications**, particularly brain-machine interfaces and neural prosthetics. The human brain operates on roughly 20 W of power; implants or wearables that interface with the brain must likewise be extremely energy-efficient and preferably real-time. Neuromorphic chips, by design, meet these criteria, making them ideal candidates for *in situ* neural signal processing or prosthetic control. For example, a neuromorphic processor could sit on a headset or implanted device, decoding neural signals from EEG or neural probes on the fly, using milliwatts of power – something not feasible with power-hungry GPUs. Recent reviews highlight neuromorphic algorithms for brain implants that could enable closed-loop systems for treating neurological conditions. Applications include seizure detection in epilepsy (where a spiking neural network could detect the onset of a seizure from neural data and trigger a stimulus to prevent it) and brain-controlled prosthetic limbs (where a neuromorphic decoder interprets motor cortex spikes to drive a robotic arm).

By 2024, some initial demonstrations have been made. For instance, researchers built ultra-low-power SNN-based classifiers that can detect epileptic seizures from intracranial EEG signals in real time, running on neuromorphic hardware with sub-milliwatt power consumption – a crucial step toward implantable seizure suppression devices. Another group implemented an SNN on a neuromorphic chip to restore a rudimentary sense of touch in a prosthetic hand, by encoding tactile sensor inputs into spikes and stimulating nerves accordingly. These examples are early, but they show how neuromorphic systems can interface with biological neural systems more naturally than conventional computers. The event-driven operation of neuromorphic chips is well-suited to processing spiking activity from the body, and their energy efficiency addresses the battery and heat constraints of implants.

The opportunities in medicine extend beyond implants: neuromorphic sensors and processors could be used for remote health monitoring, smart prosthetics, and even medical diagnostics where power and latency are critical (for example, analyzing signals in a portable brain scanner or running AI algorithms in a hearing aid). Yet, challenges or gaps remain. One gap is the maturity of the technology – regulatory approval and reliability for medical devices require robust hardware and many hours of testing, and neuromorphic systems are still largely in the research phase. Another challenge is the need for customization: neural data is complex and patient-specific, so neuromorphic algorithms must be adaptable. Techniques like on-chip learning (through local rules or few-shot learning algorithms) will be vital so that a neuromorphic implant can tune itself to an individual's neural signatures over time. Encouragingly, the trend in 2019–2024 toward online learning rules and closed-loop demonstration is directly in line with these needs. In summary, healthcare could be one of the first domains where neuromorphic computing has a tangible real-world impact, potentially improving quality of life for patients via brain-inspired, energy-efficient technology.

5.3 Scientific and Industrial Applications

Neuromorphic computing also offers opportunities in scientific research and industry, especially for applications where real-time data processing and low energy footprint are paramount. One example is in the realm of smart sensors and IoT (Internet of Things). Neuromorphic processors can be integrated with sensors (vision, auditory, olfactory, etc.) to create intelligent sensors that preprocess and interpret data on the edge. For instance, combining event-based vision cameras with neuromorphic chips yields a completely event-driven vision system that can detect objects or motion with minimal latency and power – useful for drones, mobile robots, or surveillance devices. In the early 2020s, some prototypes of neuromorphic vision systems for robotic platforms were demonstrated, where an event camera feeds directly into a spiking network running on Loihi or SpiNNaker to perform obstacle avoidance or target tracking in real-time. These systems consumed far less power than a traditional camera plus GPU setup and responded faster, since they didn't need to wait for frames and could react to each event as it happened.

Another scientific application is in **computational neuroscience and brain simulation**, which we touched on earlier. Neuromorphic hardware allows researchers to experiment with models of neural circuits at speeds comparable to biological real time or faster, which could aid hypothesis testing in neuroscience. It

also finds use in physics and network science: spiking networks have been used as analog solvers for optimization problems (e.g., solving constraint satisfaction or graph coloring by exploiting network dynamics to settle into solutions). An example is mapping a difficult optimization problem onto a network of spiking neurons such that the network's low-energy states correspond to good solutions of the problem; neuromorphic chips can then find solutions using little energy via their natural dynamics.

In industrial contexts, neuromorphic chips might be deployed in scenarios where power is at a premium – for example, satellites or remote sensors that run on solar power, or large-scale data centers looking to reduce energy costs for specific workloads. While general-purpose CPUs and GPUs still dominate, neuromorphic accelerators could carve out niches. One such niche could be **real-time control systems** (in manufacturing or automobiles) that require fast reflexes; a neuromorphic controller can process sensor inputs and output control signals with microsecond latencies. Indeed, SpiNNaker-2's team pointed out potential uses in automotive AI and tactile internet (haptic feedback systems with tight latency constraints). These are areas where even milliseconds matter and where power is limited (e.g., a self-driving car or a drone).

The gap to overcome for broader industrial adoption is largely one of software and familiarity. Most engineers and developers are versed in programming for von Neumann machines and using frameworks like TensorFlow for AI – programming a spiking neural network on a neuromorphic chip is a very different paradigm. As of 2024, the ecosystem for neuromorphic software is still maturing. Efforts like Intel's Lava framework (an open-source software for Loihi) and community-driven tools like PyNN, Brian2, or Nengo provide higher-level interfaces, but they are not yet as seamless or widely adopted as standard AI tools. Bridging this gap – by developing better compilers, libraries, and perhaps middleware that can translate parts of deep learning models to spiking equivalents – is critical for neuromorphic computing to find widespread use in industry. There is active work in creating benchmarking suites (e.g., NeuroBench) and standards for comparing neuromorphic solutions to traditional ones in application-specific contexts. If neuromorphic computing can demonstrate a clear advantage on certain tasks (like ultra-low-power sensor analytics or fast control loops) in a way that's accessible to engineers, it will secure its place in the toolkit for future smart systems.

5.4 Power-Efficiency and Scaling Challenges

Energy efficiency is the flagship advantage of neuromorphic computing. Requiring only picojoules or nano-joules per spike operation, neuromorphic chips can in principle outperform CPUs/GPUs by orders of magnitude in terms of computations per watt. This advantage has been repeatedly demonstrated in research settings – e.g., Loihi solving a constraint satisfaction problem $1000\times$ more efficiently than a CPU. However, this comes with a **trade-off**: the efficiency is best realized on problems that map well to the architecture (event-driven, sparse, parallelizable problems). If one tries to use a neuromorphic chip like a drop-in replacement for a GPU on tasks like dense matrix multiplication or large-scale number crunching, it may not fare well due to lower numerical precision and communication overhead.

A key challenge is **scaling**: how to maintain efficiency as we scale neuromorphic systems to larger sizes or broader tasks. Biological brains scale by having enormous numbers of relatively slow, low-power units operating in parallel – neuromorphic systems attempting to scale up face issues of routing millions of spikes (communication overhead can become significant), manufacturing variations (especially for analog components), and simply managing/programming very large networks. For digital chips like Loihi or TrueNorth, scaling up means adding more cores and interconnect, which at some point runs into chip area and power limits (though still far better than GPUs for equivalent neurons). For novel devices like memristors, integrating millions or billions of devices reliably is non-trivial and yield issues can arise.

Another challenge is that **efficiency alone doesn't guarantee accuracy**. A neuromorphic chip might be extremely efficient but if it cannot achieve the same accuracy or result quality as a more power-hungry device, its utility is limited. Throughout 2019–2024, we have seen this gap narrow – surrogate gradient methods have allowed SNNs to reach accuracy closer to ANNs, and certain tasks (especially those involving spatiotemporal data) where SNNs even have an edge. But generally, digital accelerators for ANNs (like TPUs) have also improved efficiency, and for many tasks they remain the easier path to high accuracy. Thus, neuromorphic computing must keep pushing the envelope not just on efficiency but on capability, to justify itself for more than niche uses.

In terms of power-efficiency metrics, a notable development is the emphasis on **event-driven benchmarks**. Traditional FLOPS/Watt is not quite applicable to spiking systems; instead, metrics like "energy per inference on dataset X" or "operations per joule for task Y" are considered. For example, one might report that a neuromorphic system can classify a DVS (event camera) gesture with 1 mJ of energy versus 100 mJ on a GPU – a compelling number. Ensuring these comparisons are fair and that neuromorphic hardware is tested on problems that play to its strengths is crucial. Initiatives in the neuromorphic community to standardize benchmarks (like IBM's NSERC or EU's NEUROTECH consortium guidelines) aim to track progress on both energy and performance.

In summary, the opportunity is clear: if neuromorphic computing can continue to scale and improve, it offers a path to sustainable computing at a time when Moore's Law is slowing and energy concerns are paramount. The gap to mind is ensuring that as energy efficiency is realized, we do not sacrifice the generality or accuracy needed for real-world applications. Closing this gap likely requires co-design of hardware and algorithms – tailoring neuromorphic substrates to support the computational primitives most useful for AI and conversely developing algorithms that can harness the hardware's strengths. The 2019–2024 period has already shown the benefits of such co-design (as seen with Loihi's features being used by new algorithms and new training methods emerging partly motivated by hardware constraints). This synergy must continue for neuromorphic computing to truly deliver on its promise of "more with less".

5.5 Integration with Conventional Computing

A practical consideration as neuromorphic technology matures is how to integrate it into existing computing systems and workflows. It's unlikely that neuromorphic chips will completely replace CPUs or GPUs; instead, they will function as **accelerators** or specialized co-processors for certain tasks, at least in the near term. Therefore, making it easy to offload computations to a neuromorphic device and get results back (much as one does with a GPU today for deep learning inference) is important.

One challenge is the communication interface: neuromorphic hardware speaks in spikes, conventional hardware in binary numbers. Bridging this involves software that can translate data into spike events and vice versa. For instance, if using a neuromorphic accelerator for a segment of a signal processing pipeline, one needs an *encoder* to convert real-valued signals into spike trains (this could be as simple as a Poisson encoder or as complex as a sensory front-end model), and a *decoder* to interpret the spiking output back into a usable form (like a class label or a control command). Developing efficient and standardized encoding/decoding schemes is an active area of research. Some schemes try to preserve information with minimal spikes (to keep energy low) while maintaining accuracy.

Another aspect of integration is **software integration**. Ideally, a machine learning engineer should be able to use a neuromorphic accelerator without needing deep expertise in spiking networks. This is where software frameworks come in: for example, there are efforts to allow training a network in PyTorch or TensorFlow and then automatically convert and deploy it to neuromorphic hardware (using tools that translate the trained ANN to an SNN and map it onto the chip). While conversion methods exist (especially for simple rate-coded networks), a fully seamless pipeline is not yet there. In 2024, Intel's Lava aimed to provide an open API where users can define networks and run them on Loihi hardware or in simulation similarly to how they'd use any AI accelerator. Ensuring such frameworks continue to develop will be key to wider adoption.

From a system architecture perspective, incorporating neuromorphic chips into computers or devices raises questions: Should they sit on the periphery near sensors? On an IoT node? Or as a card in a server? Different use cases lead to different integration strategies. For edge devices like smartphones or wearables, a neuromorphic chip could be embedded to handle always-on tasks like keyword spotting or anomaly detection in sensor data, waking up the main processor only when necessary – this is analogous to how some phones have dedicated DSPs for low-power audio processing. In cloud or HPC contexts, neuromorphic boards could be plugged in to handle particular workloads (for example, maybe handling spiking recurrent networks for certain types of simulation or running specific event-based data streams). We have already seen neuromorphic systems at some national labs exploring things like combinatorial optimization or large-scale neural simulations, often linking multiple neuromorphic boards together for more capacity.

The opportunity in integration is that neuromorphic computing doesn't have to operate in isolation; it can augment classical computing. But the gap is largely in the *interface*: both the technical interface (data formats, protocols, programming models) and the human interface (skills and tools for developers). Overcoming this will likely involve standardization: as the field coalesces, we might see standard neuron models or file formats for SNNs, analogous to how today's deep learning has standardized on certain layer types and model exchange formats. Already, initiatives like the Neuroscience-inspired Architecture (NIA) roadmap emphasize interoperability between neuromorphic and traditional systems.

In conclusion, integration is a critical step to ensure neuromorphic innovations transition from lab demos to deployed technologies. The 2019–2024 period laid groundwork by expanding software frameworks and demonstrating co-processor style usage of neuromorphic chips. The next steps will involve refining these interfaces and proving clear use-cases where a neuromorphic accelerator plugged into a conventional system provides tangible benefits in real-world applications.

6 Conclusion

Neuromorphic computing has made remarkable strides from 2019 to 2024, evolving from a collection of intriguing prototypes into a more cohesive field with demonstrated advantages in efficiency and new functionality. On the **hardware side**, we now have a spectrum of neuromorphic platforms: programmable digital chips like Loihi and SpiNNaker that can implement large spiking networks, analog and mixed-signal systems leveraging memristors or capacitive circuits for in-memory computing, spintronic and photonic devices pushing the boundaries of speed and parallelism, and emerging nanomaterial-based devices offering unprecedented integration density. Each of these approaches contributes pieces to the puzzle of brain-like computation, and ongoing research is actively exploring how to combine them (for example, hybrid systems where conventional digital logic orchestrates an ensemble of analog nanoscale devices). The hardware gains have been complemented by **algorithmic advances** – today's SNNs are far more trainable and capable than those of just a few years ago. Techniques like surrogate gradient descent and e-prop have enabled SNNs to learn complex tasks, reducing the accuracy gap with traditional neural networks while retaining the temporal processing benefits of spiking dynamics. We have also developed a better understanding of how to use spikes effectively (e.g. through temporal coding) and how to let networks learn and adapt in real-time, which could be game-changers for autonomous systems and continual learning applications.

This review also highlights that neuromorphic computing is not a monolithic technology but a **multi-faceted paradigm** – its value often appears in specialized contexts. For instance, if an application demands real-time responsiveness to streams of events under tight energy constraints (like a medical implant or an autonomous drone), neuromorphic solutions have shown they can excel. Conversely, for tasks requiring massive number crunching with extreme precision, conventional digital accelerators still hold sway. The path forward for neuromorphic computing is to capitalize on its strengths: harnessing event-driven parallelism, low-power operation, and on-chip learning to enable functionalities that would be impractical otherwise. In doing so, neuromorphic engineers will continue to collaborate with neuroscientists (to inspire new architectures and rules), material scientists (to realize new devices), and computer scientists (to create better software and integration methods).

There remain key challenges to address. These include improving the scalability and robustness of neuromorphic devices (ensuring that efficiency gains hold at large scales and under variability), developing user-friendly programming models (so that a wider community can adopt these technologies), and identifying "killer applications" that clearly demonstrate neuromorphic superiority. Encouragingly, the trend of the last five years has been positive on all these fronts: energy efficiency metrics have improved, algorithms are more sophisticated, and pilot applications in areas like sensing and prosthetics have validated core assumptions. Neuromorphic computing is steadily transitioning from a research curiosity to a practical technology.

In conclusion, the period of 2019–2024 has solidified neuromorphic computing's promise as a cornerstone for the future of computing in an era where we are constrained by energy and looking for intelligent, adaptive systems. It is unlikely to replace conventional computing wholesale – instead, it will augment and enrich it.

By continuing to learn from the ultimate computing reference (the brain) and by integrating those lessons into both hardware and software, neuromorphic computing is poised to unlock new horizons in computing capability. The next few years will be critical in moving from promising demonstrations to scalable systems working in the wild. The groundwork laid in this period gives ample reason for optimism that neuromorphic ideas will play a significant role in shaping more efficient, intelligent technologies that align with the needs of our data-driven, energy-conscious society.

7 References

- Davies et al., 2021 Mike Davies et al., "Advancing Neuromorphic Computing With Loihi: A Survey of Results and Outlook," Proc. IEEE, vol. 109, no. 5, pp. 911–934 (2021). DOI: 10.1109/JPROC.2021.3067593
- Merolla et al., 2014 Paul A. Merolla et al., "A million spiking-neuron integrated circuit with a scalable communication network and interface," Science, 345(6197):668–673 (2014). DOI: 10.1126/science.1254642
- Mayr et al., 2019 Christian Mayr, Sebastian Höppner, Steve Furber, "SpiNNaker 2: A 10 Million Core Processor System for Brain Simulation and Machine Learning," arXiv:1911.02385 [cs.ET] (2019). DOI: 10.48550/arXiv.1911.02385
- Xiao et al., 2024 Yike Xiao et al., "Recent Progress in Neuromorphic Computing from Memristive Devices to Neuromorphic Chips," Adv. Devices & Instrum., vol. 5, art. 0044 (2024). DOI: 10.34133/adi.0044
- Marrows et al., 2024 Christopher H. Marrows et al., "Neuromorphic computing with spintronics," npj Spintronics, 2:12 (2024). DOI: 10.1038/s44306-024-00019-2
- Li et al., 2024 Renjie Li et al., "Photonics for Neuromorphic Computing: Fundamentals, Devices, and Opportunities," (Adv. Materials, accepted 2024). Preprint arXiv:2311.09767. DOI: 10.1002/adma.202312825
- Choi et al., 2025 Yunseok Choi et al., "Advanced AI computing enabled by 2D material-based neuromorphic devices," npj Unconventional Computing, 2:8 (2025). DOI: 10.1038/s44335-025-00023-7
- Neftci et al., 2019 Emre O. Neftci, Hesham Mostafa, Friedemann Zenke, "Surrogate Gradient Learning in Spiking Neural Networks," IEEE Trans. Neural Netw. Learn. Syst., 30(7): 1334–1347 (2019). arXiv:1901.09948. DOI: 10.48550/arXiv.1901.09948
- Bellec et al., 2020 Guillaume Bellec et al., "A solution to the learning dilemma for recurrent networks of spiking neurons: e-prop," Nature Communications, 11:3625 (2020). DOI: 10.1038/s41467-020-17236-y
- Mostafa, 2018 Hesham Mostafa, "Supervised Learning Based on Temporal Coding in Spiking Neural Networks," IEEE Trans. Neural Netw. Learn. Syst., 29(7): 3227–3235 (2018). DOI: 10.1109/TNNLS.2017.2726060
- Roy et al., 2019 Kaushik Roy, Akhilesh Jaiswal, Priyadarshini Panda, "Towards spike-based machine intelligence with neuromorphic computing," Nature, 575(7784): 607–617 (2019). DOI: 10.1038/s41586-019-1677-2

• Pawlak & Howard, 2025 – Wiktoria A. Pawlak, Newton Howard, "Neuromorphic algorithms for brain implants: a review," Frontiers in Neuroscience, 19:1570104 (2025). DOI: 10.3389/fnins.2025.1570104