

Article

Not peer-reviewed version

---

# Synergies and Divergences Between Spiking Neural Networks and Large Language Models

---

Zhen Huan , Heng Xue <sup>\*</sup> , Zhen Li

Posted Date: 7 July 2025

doi: 10.20944/preprints202507.0525.v1

Keywords: spiking neural networks; neuromorphic computing; energy efficiency; surrogate gradient; STDP; event-driven computation; ANN-to-SNN conversion; temporal coding; low-power AI; bio-inspired learning



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

*Article*

# Synergies and Divergences Between Spiking Neural Networks and Large Language Models

Zhen Huan <sup>1</sup>, Heng Xue <sup>1,\*</sup> and Zhen Li <sup>1,2</sup>

<sup>1</sup> Department of Computer Science and Technology, Nanjing University, China

<sup>2</sup> School of Information Science and Technology, Shanghai Jiao Tong University, China

\* Correspondence: heng.xue@nju.edu.cn

## Abstract

Spiking Neural Networks (SNNs) constitute a biologically inspired class of artificial neural networks that encode and process information using discrete spike events over time, mimicking the fundamental communication mechanism of the brain. In recent years, while deep learning and large language models (LLMs) have achieved remarkable breakthroughs across a wide spectrum of tasks, SNNs have garnered renewed attention for their unique advantages in energy efficiency, temporal processing, and event-driven computation. This survey provides a comprehensive and detailed examination of SNNs within the contemporary context shaped by the dominance of LLMs and traditional deep neural networks. We begin by elucidating the fundamental concepts underpinning SNNs, including diverse neuron models, spike encoding schemes, synaptic plasticity mechanisms, and canonical network architectures. The discussion then progresses to an in-depth analysis of training methodologies, highlighting the inherent challenges due to the non-differentiable nature of spikes and surveying state-of-the-art approaches such as surrogate gradient learning, biologically inspired local plasticity rules, and ANN-to-SNN conversion techniques. A thorough overview of neuromorphic hardware platforms is presented, showcasing how their event-driven, massively parallel designs enable orders-of-magnitude improvements in energy consumption compared to conventional von Neumann architectures that support LLMs. We also explore the breadth of practical applications where SNNs demonstrate distinct advantages, ranging from real-time sensory processing and robotics to brain-machine interfaces, as well as emerging hybrid models that integrate spiking dynamics with transformer-based language models. Finally, we identify and discuss critical open challenges and promising future directions, including scalable training algorithms, hardware-software co-design, standardization efforts, and interdisciplinary collaborations that are crucial to advancing the field. By situating SNNs in the era of large language models, this survey aims to provide a holistic understanding of their current state, synergistic potential, and transformative prospects, ultimately contributing to the development of more efficient, adaptive, and biologically plausible artificial intelligence systems.

**Keywords:** spiking neural networks; neuromorphic computing; energy efficiency; surrogate gradient; STDP; event-driven computation; ANN-to-SNN conversion; temporal coding; low-power AI; bio-inspired learning

## 1. Introduction

Spiking Neural Networks (SNNs) have emerged as a biologically inspired paradigm of artificial neural computation, aiming to closely mimic the dynamics and information processing mechanisms of the human brain [1]. Unlike traditional artificial neural networks (ANNs) which operate on continuous-valued activations and rely heavily on synchronous computations, SNNs leverage discrete spike events and temporal coding to process information in a sparse, event-driven, and highly efficient manner. This fundamental difference not only offers a more faithful model of neural processing but also presents unique opportunities for developing energy-efficient neuromorphic hardware and

advancing our understanding of cognitive functions. The concept of SNNs is rooted in neuroscience, where neurons communicate through discrete electrical impulses known as action potentials or spikes [2]. The timing and patterns of these spikes encode and transmit information, enabling complex sensory processing, motor control, and cognitive functions in biological systems [3]. Inspired by these biological principles, SNNs incorporate temporal dynamics, membrane potentials, and threshold-based firing mechanisms to replicate the spike-driven communication between neurons [4]. Over the past decades, research in SNNs has advanced steadily, focusing on developing effective neuron models, learning algorithms, and hardware implementations that harness the advantages of spiking computation. However, the landscape of artificial intelligence and neural computation has witnessed a dramatic transformation with the advent and widespread success of large language models (LLMs). These models, typified by architectures such as Transformers, have revolutionized natural language processing (NLP), computer vision, and multimodal learning through their remarkable ability to scale and learn from massive datasets. LLMs have demonstrated unprecedented capabilities in generating coherent text, understanding context, and performing a wide range of tasks with minimal fine-tuning, setting new benchmarks for AI performance. Their dominance has naturally shifted much of the research focus and industrial applications toward ANN-based deep learning approaches, which are well-suited to the highly parallel and dense matrix operations that modern hardware accelerates. In this new era dominated by LLMs and deep learning, it becomes crucial to revisit and re-evaluate the role and potential of SNNs [5]. While ANNs and LLMs excel in domains demanding vast amounts of data and computational power, they often suffer from inefficiencies related to energy consumption, interpretability, and biological plausibility. SNNs, with their sparse, asynchronous, and event-driven processing, offer a complementary approach that could address some of these limitations, especially in resource-constrained environments such as edge computing, robotics, and neuromorphic platforms [6]. Moreover, the integration of temporal dynamics and spike-based learning can potentially provide richer representations and improved robustness, qualities that remain challenging for conventional deep learning models. Despite these promising attributes, SNNs face significant challenges in achieving the same level of performance and scalability as LLMs [7]. The lack of standardized and effective training algorithms, difficulties in backpropagating errors through discrete spike events, and limited availability of large-scale spiking datasets have hindered the widespread adoption of SNNs. Furthermore, the gap between neuroscientific realism and computational practicality continues to fuel debates on the best ways to model spiking neurons and their networks. Recent advances in surrogate gradient methods, spike-based backpropagation, and hybrid models combining spiking and non-spiking neurons have started to bridge this gap, suggesting new directions for research and application. This survey aims to provide a comprehensive overview of Spiking Neural Networks in the context of the current AI paradigm shaped by large language models. We explore the historical development of SNNs, their fundamental principles, and biological inspirations, alongside a detailed examination of state-of-the-art architectures, learning techniques, and neuromorphic hardware implementations [8]. Additionally, we analyze the interplay between SNNs and LLMs, investigating how spiking models can complement or integrate with large-scale transformer architectures and other deep learning frameworks [9]. Through this lens, we assess the strengths, limitations, and future prospects of SNNs as a viable and valuable approach within the broader AI ecosystem [10]. By bridging the worlds of biologically inspired spiking computation and data-driven large language models, this survey endeavors to inspire new research that harnesses the best of both paradigms [11]. As AI continues to evolve, understanding and leveraging the unique capabilities of SNNs in tandem with the powerful representational capacities of LLMs will be essential for creating intelligent systems that are not only effective but also efficient, interpretable, and adaptable across diverse applications and platforms.

## 2. Background and Motivation

The study of Spiking Neural Networks (SNNs) traces its origins to early computational neuroscience and the quest to model biological neural processes with greater fidelity than traditional artificial

neural networks (ANNs) [12]. Unlike ANNs, which typically use continuous activation functions and operate in discrete time steps, SNNs employ discrete spikes—binary events occurring at precise points in time—to communicate information. This event-driven mechanism aligns closely with how real neurons function, firing action potentials when their membrane potentials reach certain thresholds. The temporal dimension of spiking activity introduces a rich dynamic for encoding and processing information, encompassing not only spike counts but also spike timing and patterns [13]. The motivation behind SNNs is multifold. First, biological plausibility: capturing the essence of neural computation observed in living brains holds promise for understanding cognition and neural coding principles [14]. Second, energy efficiency: the sparse and asynchronous nature of spike-based communication can drastically reduce computational overhead and power consumption, which is critical for embedded systems, wearable devices, and autonomous agents operating in energy-constrained environments. Third, robustness and adaptability: SNNs have the potential to exploit temporal features of data more naturally and adapt dynamically to changing inputs through plasticity mechanisms inspired by synaptic learning rules in the brain [15]. Historically, the main challenge for SNNs has been developing efficient and scalable training algorithms [16]. The discrete, non-differentiable nature of spike events renders the application of conventional gradient-based methods difficult [17]. Early approaches relied on biologically inspired learning rules such as Spike-Timing Dependent Plasticity (STDP), which adjusts synaptic weights based on the relative timing of pre- and post-synaptic spikes. While STDP and similar rules provide a foundation for unsupervised and reinforcement learning in SNNs, they have limited applicability in complex supervised learning tasks that require optimizing large networks with millions or billions of parameters [18]. The rise of deep learning and particularly the dominance of large language models (LLMs) have brought new perspectives to neural computation. LLMs utilize dense, layered architectures such as Transformers to process vast corpora of data, achieving exceptional results in language understanding, generation, and reasoning. Their success has been fueled by advances in algorithmic design, massive parallelization on GPUs and TPUs, and availability of extensive datasets. However, the computation in LLMs is inherently synchronous, power-hungry, and lacks direct temporal dynamics analogous to biological spiking neurons. This contrast sets the stage for re-examining SNNs under the lens of modern AI research [19]. The surge of interest in neuromorphic hardware—devices designed to mimic the architecture and dynamics of biological brains—reflects a growing demand for AI systems that combine efficiency, adaptability, and real-time processing [20]. SNNs are ideally suited for deployment on such platforms due to their event-driven nature. Additionally, integrating spiking mechanisms into hybrid models or leveraging temporal coding strategies may enhance the expressiveness and generalization capabilities of AI models beyond what is achievable with traditional LLM architectures alone [21]. In this section, we provide an in-depth background on the fundamental concepts underlying SNNs, tracing their evolution from biological neurons to computational models [22]. We discuss key neuron models such as the Leaky Integrate-and-Fire (LIF) and Hodgkin-Huxley models, highlight classical learning rules, and outline the main challenges that have limited their practical adoption [23]. Furthermore, we contextualize these developments within the current AI landscape dominated by LLMs, emphasizing why revisiting SNNs now is both timely and necessary. Ultimately, understanding this background is critical for appreciating the motivations behind ongoing research efforts aimed at bridging the gap between SNNs and deep learning, and for identifying potential synergies that can drive the next generation of intelligent systems. This foundation also sets the stage for subsequent sections, which delve into contemporary architectures, training methodologies, and hardware platforms that enable SNNs to coexist and potentially integrate with the large-scale models shaping modern AI.

### 3. Fundamental Concepts of Spiking Neural Networks

Spiking Neural Networks (SNNs) represent a paradigm shift in artificial neural computation, emphasizing temporal dynamics and discrete event-based communication rather than continuous activation levels [24]. This section provides a detailed exploration of the core concepts that underpin SNNs,



including neuron models, spike encoding schemes, synaptic dynamics, and network architectures. Understanding these fundamentals is essential for grasping the unique capabilities and challenges associated with SNNs, especially in the context of their evolving role alongside large language models (LLMs) [25].

### 3.1. Biological Inspiration and Neuron Models

The foundation of SNNs lies in the biological neuron, which transmits information via action potentials or spikes [26]. When a neuron's membrane potential crosses a critical threshold, it emits a spike that propagates to connected neurons through synapses. This discrete signaling mechanism contrasts with the continuous activation functions used in conventional ANNs and enables temporal coding of information [27]. To model this behavior, various spiking neuron models have been proposed, balancing biological realism with computational tractability [28]. The most commonly used models include:

- **Leaky Integrate-and-Fire (LIF) Model:** A simplified yet effective neuron model where the membrane potential integrates incoming spikes with leakage over time [29]. When the potential exceeds a threshold, the neuron fires a spike and resets its potential. Its mathematical simplicity has made it a staple in SNN research and hardware implementations.
- **Hodgkin-Huxley Model:** A biophysically detailed model that describes ion channel dynamics responsible for action potential generation [30]. Although highly accurate, its complexity limits its use primarily to neuroscientific simulations rather than large-scale SNNs.
- **Izhikevich Model:** Combines biological plausibility and computational efficiency by approximating various spiking and bursting patterns observed in real neurons [31]. It is widely used for capturing diverse neuronal dynamics with reduced computational cost.

### 3.2. Spike Encoding and Temporal Coding

Information in SNNs is represented by the timing and patterns of spikes rather than static values [32]. Encoding continuous or high-dimensional data into spike trains is a critical step, and several strategies have been developed:

- **Rate Coding:** Represents information by the firing rate (spikes per unit time) of neurons, analogous to the average activity level in traditional ANNs [33].
- **Temporal Coding:** Uses precise spike timing to convey information, such as the latency of the first spike or relative timing between spikes, enabling more compact and efficient representations.
- **Population Coding:** Distributes information across groups of neurons where the collective spiking pattern encodes stimulus attributes, enhancing robustness and expressiveness [34].

Each encoding scheme has trade-offs in terms of biological plausibility, information capacity, and computational complexity [35].

### 3.3. Synaptic Dynamics and Plasticity

Synapses in SNNs are modeled as dynamic entities that modulate the strength and timing of spike transmission between neurons. Synaptic weights govern how incoming spikes influence the postsynaptic neuron's membrane potential [36]. Beyond static weights, synaptic plasticity mechanisms enable learning by adapting these weights based on activity patterns [37]. Key plasticity rules include:

- **Spike-Timing Dependent Plasticity (STDP):** A biologically inspired rule where synaptic strength is adjusted depending on the temporal order and interval between pre- and postsynaptic spikes, reinforcing causally linked spikes.
- **Hebbian Learning:** Synapses are strengthened when pre- and postsynaptic neurons fire together, embodying the principle that "neurons that fire together wire together."
- **Homeostatic Plasticity:** Mechanisms that stabilize neural activity by adjusting synaptic strengths to maintain balanced firing rates.

These plasticity rules contribute to the network's ability to learn and adapt in both supervised and unsupervised contexts [38].

### 3.4. Network Architectures

SNN architectures range from simple feedforward networks to complex recurrent and convolutional structures [39]. Some architectures mirror those in deep learning, adapted for spiking neurons, including:

- **Feedforward SNNs:** Basic layered networks where spikes propagate forward, suitable for pattern recognition tasks.
- **Recurrent SNNs:** Incorporate feedback loops enabling temporal memory and dynamic behavior, crucial for sequence processing and temporal pattern recognition [40].
- **Convolutional SNNs:** Adapt convolutional operations to spike-based signals, enabling spatial feature extraction similar to CNNs in ANNs.
- **Reservoir Computing and Liquid State Machines:** Utilize randomly connected recurrent networks with readout layers trained separately, leveraging the dynamic states of the reservoir.

Each architecture offers distinct advantages depending on the application domain and computational constraints [41].

### 3.5. Summary

The fundamental concepts of SNNs highlight a unique computational framework that blends temporal dynamics, event-driven processing, and biologically inspired learning. These features enable energy-efficient, robust, and adaptive neural computation, setting SNNs apart from traditional ANN-based methods, including the dominant LLMs [42]. However, the increased complexity in modeling and training SNNs also introduces significant challenges that have shaped research directions [43]. Understanding these foundational elements provides a necessary context for appreciating the advances and innovations discussed in subsequent sections of this survey [44].

## 4. Training Algorithms for Spiking Neural Networks

Training Spiking Neural Networks (SNNs) effectively remains one of the most critical and challenging areas in the field, especially when contrasted with the well-established backpropagation techniques that have propelled the success of large language models (LLMs) and other deep learning architectures. This section provides a comprehensive overview of the various learning algorithms developed for SNNs, addressing their theoretical foundations, practical implementations, and current limitations [45].

### 4.1. Challenges in Training SNNs

A central difficulty in training SNNs arises from the discrete, non-differentiable nature of spike events [46]. Unlike conventional artificial neurons that produce continuous outputs amenable to gradient-based optimization, spiking neurons emit binary spikes at specific time points [47]. This discontinuity prevents straightforward application of gradient descent and backpropagation through time (BPTT), which underpin most state-of-the-art deep learning methods. Additionally, the temporal dynamics and recurrent connections common in SNNs introduce further complexity, requiring algorithms to handle time-dependent variables and long-range dependencies efficiently [48]. These challenges have led to the exploration of alternative training paradigms and adaptations of traditional learning techniques.

### 4.2. Spike-Timing Dependent Plasticity (STDP)

One of the earliest and most biologically plausible learning rules is Spike-Timing Dependent Plasticity (STDP) [49]. STDP modulates synaptic weights based on the relative timing between pre- and postsynaptic spikes: if a presynaptic spike precedes a postsynaptic spike within a certain time window,

the synapse is potentiated; if the order is reversed, the synapse is depressed [50]. This temporally asymmetric Hebbian learning rule enables networks to self-organize and detect temporal correlations in input data. Although STDP and its variants have demonstrated success in unsupervised feature learning and associative memory tasks, they are less suited for large-scale supervised learning due to their local and unsupervised nature [51]. Researchers have attempted to extend STDP by combining it with global reward signals or hybrid frameworks, yet it remains difficult to scale these approaches to complex datasets and tasks [52].

#### 4.3. Surrogate Gradient Methods

Surrogate gradient methods have recently emerged as a powerful approach to circumvent the non-differentiability of spike generation. These methods approximate the gradient of the spiking function with a smooth surrogate during backpropagation, enabling the use of gradient descent and backpropagation through time (BPTT) to train deep SNNs [53]. By replacing the hard threshold with a differentiable proxy (e.g., a sigmoid or piecewise-linear function) during the backward pass, surrogate gradients allow error signals to flow through spike events, facilitating end-to-end supervised learning [54]. This approach has enabled SNNs to achieve competitive performance on image recognition, speech processing, and reinforcement learning tasks. However, surrogate gradient methods often require careful tuning of surrogate functions and hyperparameters, and the temporal credit assignment problem remains challenging for long sequences [55].

#### 4.4. Conversion from Pretrained ANNs

Another practical strategy is ANN-to-SNN conversion, where a conventional trained artificial neural network is transformed into an equivalent SNN [56]. By mapping activation values to firing rates and adjusting thresholds, this approach leverages the performance of mature ANN models while gaining the energy efficiency and event-driven benefits of spiking computation [57]. Conversion methods have been effective in image classification tasks using convolutional networks but face challenges in temporal and sequential domains, as well as in handling temporal coding schemes beyond rate coding [58].

#### 4.5. Other Learning Paradigms

Additional approaches include reinforcement learning, evolutionary algorithms, and local learning rules tailored to specific tasks or hardware constraints [59]. These methods explore alternative avenues for training SNNs when gradient-based techniques are not feasible or efficient [60].

#### 4.6. Summary and Outlook

While no single training method currently dominates SNN learning as backpropagation does for LLMs, advances in surrogate gradients and hybrid training methods have significantly narrowed the performance gap. Continued research into scalable, biologically plausible, and hardware-friendly learning algorithms is crucial to unlocking the full potential of SNNs. Integrating insights from neuroscience, machine learning, and neuromorphic engineering will likely drive the development of novel training frameworks capable of synergizing with, or complementing, the capabilities of large language models in the future.

### 5. Neuromorphic Hardware and Computational Efficiency

A defining promise of Spiking Neural Networks (SNNs) lies in their potential to dramatically improve computational efficiency and energy consumption compared to traditional artificial neural networks (ANNs), especially large-scale models such as large language models (LLMs). This section explores the advances in neuromorphic hardware designed specifically to leverage the sparse, event-driven nature of SNNs, and discusses how such platforms contrast with conventional hardware optimized for dense matrix operations typical of LLMs [61].

### 5.1. Principles of Neuromorphic Computing

Neuromorphic computing architectures are inspired by the brain's structure and operational principles, aiming to emulate neural processing through massively parallel, distributed, and asynchronous computation [62]. These systems implement spiking neurons and synapses directly in hardware, enabling them to operate with low latency and high energy efficiency by processing information only when spikes occur, rather than continuously [63]. Key principles include:

- **Event-driven Processing:** Computation and communication happen only on spike events, reducing idle cycles and unnecessary power usage.
- **Local Memory and Computation:** Mimicking biological neurons, memory and processing units are colocated, alleviating the von Neumann bottleneck common in traditional architectures [64].
- **Massive Parallelism:** Large numbers of neurons and synapses operate simultaneously, enabling real-time processing of complex sensory data [65].

### 5.2. Notable Neuromorphic Platforms

Several neuromorphic hardware platforms have been developed to harness the advantages of SNNs:

- **IBM TrueNorth:** Featuring over one million spiking neurons and 256 million synapses, TrueNorth uses a manycore architecture with event-driven communication, achieving orders of magnitude reduction in power consumption compared to GPUs.
- **Intel Loihi:** A programmable neuromorphic chip supporting on-chip learning through spike-based plasticity rules, Loihi facilitates real-time adaptive systems and has been applied in robotics and sensory processing.
- **SpiNNaker:** Designed as a massively parallel digital computer mimicking neural architectures, SpiNNaker supports large-scale SNN simulations with high flexibility but uses traditional processors [66].
- **BrainScaleS:** An analog neuromorphic platform that exploits the physical properties of electronic circuits to emulate neural dynamics at accelerated timescales [67].

Each platform offers distinct trade-offs in programmability, scalability, precision, and power efficiency, reflecting diverse approaches to neuromorphic design [68].

### 5.3. Energy Efficiency and Performance Comparisons

Neuromorphic hardware demonstrates significant reductions in energy consumption—often by several orders of magnitude—compared to GPUs and TPUs running conventional deep learning workloads [69]. This efficiency stems from the sparse, asynchronous spike-driven computations that avoid continuous floating-point operations. However, these gains come with limitations such as lower precision, limited memory capacity, and challenges in mapping complex architectures like large transformers directly onto neuromorphic substrates. The maturity of software toolchains, training algorithms, and integration with existing AI frameworks also remains a bottleneck.

### 5.4. Challenges and Opportunities in Integrating with LLMs

Large language models require massive computational resources and benefit from highly optimized dense matrix multiplications on traditional hardware [70]. Bridging SNNs and neuromorphic computing with LLM-scale workloads is an open challenge. Promising research directions include:

- **Hybrid Architectures:** Combining spiking layers with conventional ANN components to leverage efficiency without sacrificing performance [71].
- **Event-driven Transformers:** Developing spike-based approximations of transformer architectures to exploit temporal sparsity.
- **Algorithm-Hardware Co-design:** Tailoring training algorithms and network architectures specifically for neuromorphic hardware capabilities [72].



### 5.5. Summary

Neuromorphic hardware exemplifies the practical advantages of SNNs in terms of energy-efficient, low-latency computation. While current systems have yet to rival the raw performance and scalability of hardware used for LLMs, ongoing advances in device technology, architecture, and algorithmic integration point toward a future where spiking and conventional AI models coexist and complement each other. This synergy could unlock novel applications demanding real-time, adaptive, and low-power intelligence beyond the reach of today's deep learning frameworks [73].

## 6. Applications and Emerging Trends

As Spiking Neural Networks (SNNs) continue to evolve alongside the rapid advancements in large language models (LLMs) and deep learning, their unique properties are driving a range of novel applications and inspiring emerging research trends [74]. This section surveys key application domains where SNNs have demonstrated promise, as well as cutting-edge directions that explore the integration and hybridization of spiking and non-spiking architectures.

### 6.1. Energy-Efficient and Real-Time Systems

One of the most immediate and impactful application areas for SNNs is in energy-constrained and real-time environments [75]. The sparse, event-driven nature of spike-based computation makes SNNs particularly suited for deployment on edge devices, autonomous robots, wearable sensors, and Internet of Things (IoT) systems, where low power consumption and fast response times are critical [76]. Examples include:

- **Neuromorphic Vision and Auditory Processing:** SNNs excel at processing event-based sensory inputs such as those from Dynamic Vision Sensors (DVS) or silicon cochleas, enabling ultra-low-latency object detection, gesture recognition, and auditory scene analysis.
- **Robotics and Control:** Real-time adaptive control systems leverage SNNs for processing sensorimotor feedback with minimal energy, supporting navigation, manipulation, and interaction in dynamic environments [77].

### 6.2. Brain-Machine Interfaces and Neuroprosthetics

The close alignment of SNNs with biological neural dynamics has fostered their application in brain-machine interfaces (BMIs) and neuroprosthetic devices. By interpreting neural spike patterns and generating bio-compatible signals, SNNs can facilitate communication between the nervous system and external devices, advancing rehabilitation and augmentation technologies [78].

### 6.3. Hybrid Models and Cross-Paradigm Integration

Recent research explores hybrid architectures that combine the strengths of SNNs and traditional ANNs or LLMs [79]. Such approaches seek to leverage the representational power and scalability of deep learning with the temporal precision and efficiency of spiking neurons [80]. Examples include:

- **Spiking Transformers:** Adaptations of transformer architectures employing spiking mechanisms to incorporate temporal coding and event-driven processing [81].
- **Neuromorphic Preprocessing:** Using SNNs as front-end feature extractors or encoders for subsequent processing by LLMs or deep neural networks, enabling efficient sensory data compression [82].
- **Joint Training Frameworks:** Developing algorithms that enable end-to-end training of mixed spiking and non-spiking layers for complex tasks [83].

### 6.4. Learning and Adaptation in Dynamic Environments

SNNs' inherent capacity for temporal coding and biologically inspired plasticity mechanisms makes them well-suited for continual learning and adaptation in non-stationary environments. Emerging trends focus on leveraging SNNs for lifelong learning, robust anomaly detection, and flexible task

switching without catastrophic forgetting, challenges that remain difficult for standard deep learning models.

### 6.5. Theoretical and Neuroscientific Insights

Beyond engineering applications, SNN research continues to provide valuable insights into neural coding, cognitive function, and brain-inspired computation. Modeling higher-level cognitive processes such as attention, memory, and decision-making using spiking frameworks bridges the gap between neuroscience and artificial intelligence, informing both fields.

### 6.6. Summary

The landscape of SNN applications is expanding rapidly, fueled by advances in neuromorphic hardware, training algorithms, and hybrid modeling approaches. While large language models dominate many areas of AI, SNNs carve out essential niches where energy efficiency, temporal processing, and biological plausibility are paramount [84]. The emerging convergence of spiking and non-spiking paradigms heralds a future of more versatile, adaptive, and efficient intelligent systems capable of addressing a broader spectrum of real-world challenges.

## 7. Future Directions and Open Challenges

Despite significant progress, Spiking Neural Networks (SNNs) face a series of open challenges and research opportunities that must be addressed to fully realize their potential, especially in an era dominated by large language models (LLMs) and deep learning. This section outlines key future directions and unresolved issues that are critical for advancing SNN theory, applications, and integration with contemporary AI systems [85].

### 7.1. Scalable and Efficient Training Algorithms

Current training methods for SNNs, including surrogate gradient techniques and ANN-to-SNN conversion, have made remarkable strides but still lag behind the efficiency, scalability, and generalization capabilities of gradient-based learning in traditional deep networks [86,87]. Developing novel, biologically plausible, and hardware-aware learning algorithms that can scale to deep, recurrent, and transformer-like spiking architectures remains a pressing challenge [88]. Research into hybrid training paradigms combining local plasticity rules with global error signals, meta-learning approaches, and self-supervised learning adapted to spiking dynamics could open new pathways toward more effective learning [89].

### 7.2. Bridging the Gap with Large Language Models

Large language models represent the state-of-the-art in natural language understanding and generation but operate fundamentally differently from spiking systems. Exploring how temporal coding, spike-based representations, and event-driven computation can be integrated or hybridized with LLM architectures is an exciting frontier. Possible avenues include developing spiking approximations of transformer blocks, neuromorphic hardware accelerators tailored for sequence modeling, and algorithms that exploit spike timing to enhance interpretability and efficiency in language tasks [90].

### 7.3. Neuromorphic Hardware Co-Design

The full benefits of SNNs can only be realized through tight co-design of algorithms, architectures, and hardware [91]. Future research should focus on designing flexible, scalable neuromorphic platforms that support on-chip learning, efficient memory usage, and interoperability with existing AI ecosystems [92]. Advances in emerging device technologies such as memristors, photonic spiking neurons, and hybrid analog-digital systems may provide new opportunities for ultra-low-power, high-throughput neuromorphic computing.

#### 7.4. Standardization and Benchmarking

A significant obstacle in SNN research is the lack of standardized benchmarks, datasets, and evaluation protocols comparable to those used for deep learning and LLMs [93]. Establishing widely accepted benchmarks for classification, sequence modeling, continual learning, and real-world applications will facilitate fair comparisons, reproducibility, and community-driven progress.

#### 7.5. Interdisciplinary Collaboration

SNN research sits at the intersection of neuroscience, computer science, and electrical engineering. Strengthening interdisciplinary collaborations will accelerate the translation of biological insights into computational models and hardware, while also informing neuroscience with new computational theories and tools [94].

#### 7.6. Ethical and Societal Implications

As SNNs and neuromorphic systems mature and potentially become integral to edge computing, healthcare, and autonomous systems, it is essential to consider ethical implications related to privacy, security, and societal impact [95]. Designing transparent, interpretable, and trustworthy spiking AI systems will be an important focus moving forward [96].

#### 7.7. Summary

The future of Spiking Neural Networks is promising yet complex, requiring breakthroughs in theory, algorithms, hardware, and applications [97]. Addressing these challenges through innovative research and cross-disciplinary efforts will enable SNNs to complement and possibly transform the AI landscape dominated by large language models, leading to more efficient, adaptive, and biologically inspired intelligent systems.

### 8. Conclusion

Spiking Neural Networks (SNNs) represent a compelling and biologically inspired approach to artificial intelligence, distinguished by their event-driven, temporal coding mechanisms that promise unparalleled energy efficiency and real-time processing capabilities. In the age of large language models (LLMs) and deep learning, SNNs offer a complementary paradigm that addresses fundamental limitations of conventional neural networks, particularly in scenarios requiring low-power computation, temporal dynamics, and adaptability.

This survey has provided an extensive overview of the foundational principles of SNNs, including neuron models, spike encoding schemes, synaptic plasticity, and network architectures. We explored the diverse array of training algorithms, highlighting the challenges posed by spike non-differentiability and the innovative solutions such as surrogate gradients and ANN-to-SNN conversion. The emergence of neuromorphic hardware platforms illustrates the practical potential of SNNs for energy-efficient and scalable computation, contrasting sharply with the resource-intensive demands of LLMs.

Furthermore, we examined the broad spectrum of applications where SNNs are uniquely advantageous, from neuromorphic sensory processing and robotics to brain-machine interfaces, as well as the nascent efforts to hybridize spiking and non-spiking models. Finally, we identified critical open challenges and future research directions, emphasizing the need for scalable training methods, hardware-algorithm co-design, interdisciplinary collaboration, and ethical considerations.

As artificial intelligence continues to evolve, the convergence of SNNs and LLMs promises novel computational frameworks that blend biological plausibility with engineering efficacy. By advancing both theoretical understanding and practical implementations, the community can unlock new frontiers of intelligent systems that are more efficient, adaptive, and aligned with the complexities of natural cognition.

The journey of SNNs in the era of LLMs is just beginning, and the coming years are poised to witness exciting breakthroughs that could reshape the AI landscape.

## References

1. Shrestha, S.B.; Orchard, G. SLAYER: Spike layer error reassignment in time. In Proceedings of the Proceedings of the International Conference on Neural Information Processing Systems, 2018, pp. 1412–1421.
2. Zhang, J.; Shen, J.; Wang, Z.; Guo, Q.; Yan, R.; Pan, G.; Tang, H. SpikingMiniLM: Energy-efficient Spiking Transformer for Natural Language Understanding. *Science China Information Sciences* **2024**.
3. Gerstner, W.; Kistler, W.M. *Spiking neuron models: Single neurons, populations, plasticity*; Cambridge University Press, 2002.
4. Guo, Y.; Chen, Y.; Zhang, L.; Liu, X.; Wang, Y.; Huang, X.; Ma, Z. IM-loss: information maximization loss for spiking neural networks. *Advances in Neural Information Processing Systems* **2022**, *35*, 156–166.
5. APT Advanced Processor Technologies Research Group. SpiNNaker.
6. Sengupta, A.; Ye, Y.; Wang, R.; Liu, C.; Roy, K. Going deeper in spiking neural networks: VGG and residual architectures. *Frontiers in Neuroscience* **2019**, *13*, 95.
7. Zou, S.; Mu, Y.; Zuo, X.; Wang, S.; Cheng, L. Event-based human pose tracking by spiking spatiotemporal transformer. *arXiv preprint arXiv:2303.09681* **2023**.
8. Scellier, B.; Bengio, Y. Equilibrium propagation: Bridging the gap between energy-based models and backpropagation. *Frontiers in Computational Neuroscience* **2017**, *11*, 24.
9. Zhou, Z.; Zhu, Y.; He, C.; Wang, Y.; Yan, S.; Tian, Y.; Yuan, L. Spikformer: When spiking neural network meets transformer. In Proceedings of the Proceedings of the International Conference on Learning Representations, 2023.
10. Deng, S.; Li, Y.; Zhang, S.; Gu, S. Temporal Efficient Training of Spiking Neural Network via Gradient Re-weighting. In Proceedings of the Proceedings of the International Conference on Learning Representations, 2022.
11. de Vries, A. The growing energy footprint of artificial intelligence. *Joule* **2023**, *7*, 2191–2194.
12. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems* **2012**, *25*, 1097–1105.
13. Song, X.; Song, A.; Xiao, R.; Sun, Y. One-step Spiking Transformer with a Linear Complexity. In Proceedings of the Proceedings of the International Joint Conference on Artificial Intelligence, 2024.
14. Yao, M.; Hu, J.; Zhou, Z.; Yuan, L.; Tian, Y.; Xu, B.; Li, G. Spike-driven transformer. *Advances in Neural Information Processing Systems* **2024**, *36*, 64043–64058.
15. Desai, K.; Johnson, J. Virtex: Learning visual representations from textual annotations. In Proceedings of the Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2021, pp. 11162–11173.
16. Esser, S.K.; Merollaa, P.A.; Arthura, J.V.; Cassidy, A.S.; Appuswamy, R.; Andreopoulos, A.; Berga, D.J.; McKinstry, J.L.; Melanoa, T.; Barcha, D.R.; et al. Convolutional networks for fast energy-efficient neuromorphic computing. *Proceedings of the National Academy of Sciences of the United States of America* **2016**, *113*, 11441–11446.
17. Ioffe, S.; Szegedy, C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In Proceedings of the Proceedings of the International Conference on Machine Learning, 2015, pp. 448–456.
18. Zhu, X.; Zhao, B.; Ma, D.; Tang, H. An efficient learning algorithm for direct training deep spiking neural networks. *IEEE Transactions on Cognitive and Developmental Systems* **2022**.
19. Vigneron, A.; Martinet, J. A critical survey of STDP in Spiking Neural Networks for Pattern Recognition. In Proceedings of the Proceedings of the International Joint Conference on Neural Networks, 2020, pp. 1–9.
20. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. In Proceedings of the Proceedings of the International Conference on Learning Representations, 2015.
21. Hu, Y.; Zheng, Q.; Jiang, X.; Pan, G. Fast-SNN: Fast Spiking Neural Network by Converting Quantized ANN. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **2023**, *45*, 14546–14562.
22. Park, E.; Ahn, J.; Yoo, S. Weighted-entropy-based quantization for deep neural networks. In Proceedings of the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2017, pp. 5456–5464.
23. Ambrogio, S.; Narayanan, P.; Tsai, H.; Shelby, R.M.; Boybat, I.; di Nolfo, C.; Sidler, S.; Giordano, M.; Bodini, M.; Farinha, N.C.; et al. Equivalent-accuracy accelerated neural-network training using analogue memory. *Nature* **2018**, *558*, 60–67.
24. Berner, R.; Brandli, C.; Yang, M.; Liu, S.C.; Delbruck, T. A 240× 180 10mW 12us latency sparse-output vision sensor for mobile applications. In Proceedings of the Symposium on VLSI Circuits. IEEE, 2013, pp. C186–C187.



25. Nair, V.; Hinton, G.E. Rectified linear units improve restricted boltzmann machines. In Proceedings of the Proceedings of the International Conference on Machine Learning, 2010.
26. Krizhevsky, A.; Hinton, G. Learning multiple layers of features from tiny images. *Master's thesis, Department of Computer Science, University of Toronto* **2009**.
27. Bengio, Y.; Simard, P.; Frasconi, P. Learning long-term dependencies with gradient descent is difficult. *IEEE Transactions on Neural Networks* **1994**, *5*, 157–166.
28. Yang, S.; Ma, H.; Yu, C.; Wang, A.; Li, E.P. SDiT: Spiking Diffusion Model with Transformer. *arXiv preprint arXiv:2402.11588* **2024**.
29. Lichtsteiner, P.; Posch, C.; Delbruck, T. A  $128 \times 128$  120db 30mw asynchronous vision sensor that responds to relative intensity change. In Proceedings of the IEEE International Solid-State Circuits Conference. IEEE, 2006, pp. 2060–2069.
30. Wang, H.; Liang, X.; Li, M.; Zhang, T. RTFormer: Re-parameter TSN Spiking Transformer. *arXiv preprint arXiv:2406.14180* **2024**.
31. Eshraghian, J.K.; Ward, M.; Neftci, E.; Wang, X.; Lenz, G.; Dwivedi, G.; Bennamoun, M.; Jeong, D.S.; Lu, W.D. Training spiking neural networks using lessons from deep learning. *arXiv preprint arXiv:2109.12894* **2021**.
32. Huang, Z.; Shi, X.; Hao, Z.; Bu, T.; Ding, J.; Yu, Z.; Huang, T. Towards High-performance Spiking Transformers from ANN to SNN Conversion. In Proceedings of the Proceedings of the ACM Multimedia, 2024.
33. IBM Corporation. SyNAPSE.
34. Shen, S.; Zhao, D.; Shen, G.; Zeng, Y. TIM: An Efficient Temporal Interaction Module for Spiking Transformer. *arXiv preprint arXiv:2401.11687* **2024**.
35. Yan, J.; Liu, Q.; Zhang, M.; Feng, L.; Ma, D.; Li, H.; Pan, G. Efficient spiking neural network design via neural architecture search. *Neural Networks* **2024**, p. 106172.
36. Albanie, S. MCN-Models.
37. Maass, W. Networks of spiking neurons: the third generation of neural network models. *Neural Networks* **1997**, *10*, 1659–1671.
38. Lichtsteiner, P.; Posch, C.; Delbruck, T. A  $128 \times 128$  120 dB  $15\mu$  s Latency Asynchronous Temporal Contrast Vision Sensor. *IEEE Journal of Solid-State Circuits* **2008**, *43*, 566–576.
39. Borst, A.; Theunissen, F.E. Information theory and neural coding. *Nature Neuroscience* **1999**, *2*, 947–957.
40. Ning, Q.; Hesham, M.; Fabio, S.; Dora, S. A reconfigurable on-line learning spiking neuromorphic processor comprising 256 neurons and 128K synapses. *Frontiers in Neuroscience* **2015**, *9*, 141.
41. Intel Corporation. Intel Xeon Platinum 9282 Processor.
42. Yao, M.; Hu, J.; Hu, T.; Xu, Y.; Zhou, Z.; Tian, Y.; Xu, B.; Li, G. Spike-driven Transformer V2: Meta Spiking Neural Network Architecture Inspiring the Design of Next-generation Neuromorphic Chips. In Proceedings of the Proceedings of the International Conference on Learning Representation, 2024.
43. Guo, Y.; Zhang, Y.; Chen, Y.; Peng, W.; Liu, X.; Zhang, L.; Huang, X.; Ma, Z. Membrane potential batch normalization for spiking neural networks. In Proceedings of the Proceedings of the IEEE/CVF International Conference on Computer Vision, 2023, pp. 19420–19430.
44. Russakovsky, O.; Deng, J.; Su, H.; Krause, J.; Satheesh, S.; Ma, S.; Huang, Z.; Karpathy, A.; Khosla, A.; Bernstein, M.; et al. ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision* **2015**, *115*, 211–252. <https://doi.org/10.1007/s11263-015-0816-y>.
45. Furber, S.B.; Galluppi, F.; Temple, S.; Plana, L.A. The SpiNNaker project. *Proceedings of the IEEE* **2014**, *102*, 652–665.
46. Zhou, C.; Zhang, H.; Zhou, Z.; Yu, L.; Ma, Z.; Zhou, H.; Fan, X.; Tian, Y. Enhancing the Performance of Transformer-based Spiking Neural Networks by Improved Downsampling with Precise Gradient Backpropagation. *arXiv preprint arXiv:2305.05954* **2023**.
47. Nvidia. Nvidia V100 Tensor Core GPU.
48. Xu, B.; Geng, H.; Yin, Y.; Li, P. DISTA: Denoising Spiking Transformer with intrinsic plasticity and spatiotemporal attention. *arXiv preprint arXiv:2311.09376* **2023**.
49. Eshraghian, J.K.; Ward, M.; Neftci, E.O.; Wang, X.; Lenz, G.; Dwivedi, G.; Bennamoun, M.; Jeong, D.S.; Lu, W.D. Training spiking neural networks using lessons from deep learning. *Proceedings of the IEEE* **2023**, 111.
50. Masquelier, T.; Thorpe, S.J. Unsupervised learning of visual features through spike timing dependent plasticity. *PLoS Computational Biology* **2007**, *3*, e31.

51. Arafa, Y.; ElWazir, A.; ElKanishy, A.; Aly, Y.; Elsayed, A.; Badawy, A.H.; Chennupati, G.; Eidenbenz, S.; Santhi, N. Verified instruction-level energy consumption measurement for NVIDIA GPUs. In Proceedings of the Proceedings of the ACM International Conference on Computing Frontiers, 2020, pp. 60–70.
52. Bu, T.; Fang, W.; Ding, J.; Dai, P.; Yu, Z.; Huang, T. Optimal ANN-SNN Conversion for High-accuracy and Ultra-low-latency Spiking Neural Networks. In Proceedings of the Proceedings of the International Conference on Learning Representations, 2022.
53. LeCun, Y.; Bottou, L.; Bengio, Y.; Haffner, P. Gradient-based learning applied to document recognition. *Proceedings of the IEEE* **1998**, *86*, 2278–2324.
54. Cai, Z.; He, X.; Sun, J.; Vasconcelos, N. Deep learning with low precision by half-wave gaussian quantization. In Proceedings of the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2017, pp. 5918–5926.
55. Chowdhury, S.S.; Rathi, N.; Roy, K. One timestep is all you need: Training spiking neural networks with ultra low latency. *arXiv preprint arXiv:2110.05929* **2021**.
56. Thorpe, S.; Fize, D.; Marlot, C. Speed of processing in the human visual system. *Nature* **1996**, *381*, 520–522.
57. Brown, T.; Mann, B.; Ryder, N.; Subbiah, M.; Kaplan, J.D.; Dhariwal, P.; Neelakantan, A.; Shyam, P.; Sastry, G.; Askell, A.; et al. Language models are few-shot learners. *Advances in neural information processing systems* **2020**, *33*, 1877–1901.
58. Schneider, M.L.; Donnelly, C.A.; Russek, S.E.; Baek, B.; Pufall, M.R.; Hopkins, P.F.; Dresselhaus, P.D.; Benz, S.P.; Rippard, W.H. Ultralow power artificial synapses using nanotextured magnetic Josephson junctions. *Science Advances* **2018**, *4*, e1701329.
59. Horowitz, M. Energy table for 45nm process. *Stanford VLSI wiki* **2014**.
60. Hunsberger, E.; Eliasmith, C. Spiking deep networks with LIF neurons. *arXiv preprint arXiv:1510.08829* **2015**.
61. Kim, H.; Park, J.; Lee, C.; Kim, J.J. Improving accuracy of binary neural networks using unbalanced activation distribution. In Proceedings of the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 7862–7871.
62. Schaefer, A.T.; Margrie, T.W. Spatiotemporal representations in the olfactory system. *Trends in Neurosciences* **2007**, *30*, 92–100.
63. AMD. AMD Radeon VII GPU.
64. iniLabs. DVS 128.
65. Deng, S.; Gu, S. Optimal conversion of conventional artificial neural networks to spiking neural networks. *Proceedings of the International Conference on Learning Representations* **2021**.
66. O'Connor, P.; Neil, D.; Liu, S.C.; Delbruck, T.; Pfeiffer, M. Real-time classification and sensor fusion with a spiking deep belief network. *Frontiers in Neuroscience* **2013**, *7*, 178.
67. Indiveri, G.; Corradi, F.; Qiao, N. Neuromorphic architectures for spiking deep neural networks. In Proceedings of the Proceedings of the IEEE International Electron Devices Meeting, 2015, pp. 4–2.
68. Guo, Y.; Huang, X.; Ma, Z. Direct learning-based deep spiking neural networks: a review. *Frontiers in Neuroscience* **2023**, *17*, 1209795.
69. Collobert, R.; Weston, J.; Bottou, L.; Karlen, M.; Kavukcuoglu, K.; Kuksa, P. Natural language processing (almost) from scratch. *Journal of Machine Learning Research* **2011**, *12*, 2493–2537.
70. Van Rullen, R.; Thorpe, S.J. Rate coding versus temporal order coding: what the retinal ganglion cells tell the visual cortex. *Neural Computation* **2001**, *13*, 1255–1283.
71. Google. Cloud TPU.
72. Meng, Q.; Xiao, M.; Yan, S.; Wang, Y.; Lin, Z.; Luo, Z.Q. Training High-Performance Low-Latency Spiking Neural Networks by Differentiation on Spike Representation. In Proceedings of the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022, pp. 12444–12453.
73. Xu, Z.; Lin, M.; Liu, J.; Chen, J.; Shao, L.; Gao, Y.; Tian, Y.; Ji, R. ReCU: Reviving the dead weights in binary neural networks. In Proceedings of the Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021, pp. 5198–5208.
74. Lin, D.; Talathi, S.; Annapureddy, S. Fixed point quantization of deep convolutional networks. In Proceedings of the Proceedings of the International Conference on Machine Learning, 2016, pp. 2849–2858.
75. Li, Y.; Deng, S.; Dong, X.; Gong, R.; Gu, S. A free lunch from ANN: Towards efficient, accurate spiking neural networks calibration. In Proceedings of the Proceedings of the International Conference on Machine Learning, 2021, pp. 6316–6325.
76. Li, Y.; Dong, X.; Wang, W. Additive powers-of-two quantization: An efficient non-uniform discretization for neural networks. *Proceedings of the International Conference on Learning Representations* **2020**.

77. Li, T.; Liu, W.; Lv, C.; Xu, J.; Zhang, C.; Wu, M.; Zheng, X.; Huang, X. SpikeCLIP: A Contrastive Language-Image Pretrained Spiking Neural Network. *arXiv preprint arXiv:2310.06488* **2023**.
78. Ponulak, F.; Kasiński, A. Supervised learning in spiking neural networks with ReSuMe: sequence learning, classification, and spike shifting. *Neural Computation* **2010**, *22*, 467–510.
79. Fang, Y.; Wang, Z.; Zhang, L.; Cao, J.; Chen, H.; Xu, R. Spiking Wavelet Transformer. *arXiv preprint arXiv:2403.11138* **2024**.
80. Lian, S.; Shen, J.; Wang, Z.; Tang, H. IM-LIF: Improved Neuronal Dynamics With Attention Mechanism for Direct Training Deep Spiking Neural Network. *IEEE Transactions on Emerging Topics in Computational Intelligence* **2024**.
81. Hinton, G. The forward-forward algorithm: Some preliminary investigations. *arXiv preprint arXiv:2212.13345* **2022**.
82. Serrano-Gotarredona, T.; Linares-Barranco, B. A  $128 \times 128$  1.5% Contrast Sensitivity 0.9% FPN 3  $\mu$ s Latency 4 mW Asynchronous Frame-Free Dynamic Vision Sensor Using Transimpedance Preamplifiers. *IEEE Journal of Solid-State Circuits* **2013**, *48*, 827–838.
83. Hu, Y.; Tang, H.; Pan, G. Spiking Deep Residual Networks. *IEEE Transactions on Neural Networks and Learning Systems* **2021**, *34*, 5200–5205.
84. Bi, G.q.; Poo, M.m. Synaptic modifications in cultured hippocampal neurons: dependence on spike timing, synaptic strength, and postsynaptic cell type. *Journal of Neuroscience* **1998**, *18*, 10464–10472.
85. Bu, T.; Ding, J.; Yu, Z.; Huang, T. Optimized potential initialization for low-latency spiking neural networks. In Proceedings of the Proceedings of the AAAI Conference on Artificial Intelligence, 2022, Vol. 36, pp. 11–20.
86. Wang, Q.; Zhang, D.; Zhang, T.; Xu, B. Attention-free Spikformer: Mixing Spike Sequences with Simple Linear Transforms. *arXiv preprint arXiv:2308.02557* **2023**.
87. Zniyed, Y.; Nguyen, T.P.; et al. Enhanced network compression through tensor decompositions and pruning. *IEEE Transactions on Neural Networks and Learning Systems* **2024**.
88. Zhu, R.J.; Zhao, Q.; Zhang, T.; Deng, H.; Duan, Y.; Zhang, M.; Deng, L.J. TCJA-SNN: Temporal-channel joint attention for spiking neural networks. *arXiv preprint arXiv:2206.10177* **2022**.
89. Ding, R.; Chin, T.W.; Liu, Z.; Marculescu, D. Regularizing activation distribution for training binarized deep networks. In Proceedings of the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, pp. 11408–11417.
90. Xing, X.; Gao, B.; Zhang, Z.; Clifton, D.A.; Xiao, S.; Du, L.; Li, G.; Zhang, J. SpikeLLM: Scaling up Spiking Neural Network to Large Language Models via Saliency-based Spiking. *arXiv preprint arXiv:2407.04752* **2024**.
91. Wikipedia contributors. Tensor processing unit — Wikipedia, The Free Encyclopedia. [https://en.wikipedia.org/w/index.php?title=Tensor\\_processing\\_unit&oldid=958810965](https://en.wikipedia.org/w/index.php?title=Tensor_processing_unit&oldid=958810965), 2020. [Online; accessed 1-June-2020].
92. Liu, Z.; Shen, Z.; Savvides, M.; Cheng, K.T. Reactnet: Towards precise binary neural network with generalized activation functions. In Proceedings of the Proceedings of the European Conference on Computer Vision. Springer, 2020, pp. 143–159.
93. Wang, X.; Wu, Z.; Rong, Y.; Zhu, L.; Jiang, B.; Tang, J.; Tian, Y. SSTFormer: bridging spiking neural network and memory support transformer for frame-event based recognition. *arXiv preprint arXiv:2308.04369* **2023**.
94. O'Connor, P.; Gavves, E.; Welling, M. Training a spiking neural network with equilibrium propagation. In Proceedings of the Proceedings of the International Conference on Artificial Intelligence and Statistics, 2019, pp. 1516–1523.
95. VanRullen, R.; Thorpe, S.J. Surfing a spike wave down the ventral stream. *Vision Research* **2002**, *42*, 2593–2615.
96. Liu, Z.; Wu, B.; Luo, W.; Yang, X.; Liu, W.; Cheng, K.T. Bi-real net: Enhancing the performance of 1-bit cnns with improved representational capability and advanced training algorithm. In Proceedings of the Proceedings of the European Conference on Computer Vision, 2018, pp. 722–737.
97. Güttig, R.; Sompolinsky, H. The tempotron: a neuron that learns spike timing-based decisions. *Nature neuroscience* **2006**, *9*, 420–428.

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.