Neuromorphic Sequential Arena: A Benchmark for Neuromorphic Temporal Processing

Xinyi Chen[1], **Chenxiang Ma**[1], **Yujie Wu**[2], **Kay Chen Tan**[1,3] and **Jibin Wu**[1,2,3,*]
[1]Department of Data Science and Artificial Intelligence, The Hong Kong Polytechnic University
[2]Department of Computing, The Hong Kong Polytechnic University
[3]Research Center of Data Science and Artificial Intelligence, The Hong Kong Polytechnic University
{xinyi-97.chen, chenxiang.ma}@connect.polyu.hk, wu-yj16@tsinghua.org.cn,
{kctan, jibin.wu}@polyu.edu.hk

Abstract

Temporal processing is vital for extracting meaningful information from time-varying signals. Recent advancements in Spiking Neural Networks (SNNs) have shown immense promise in efficiently processing these signals. However, progress in this field has been impeded by the lack of effective and standardized benchmarks, which complicates the consistent measurement of technological advancements and limits the practical applicability of SNNs. To bridge this gap, we introduce the Neuromorphic Sequential Arena (NSA), a comprehensive benchmark that offers an effective, versatile, and application-oriented evaluation framework for neuromorphic temporal processing. The NSA includes seven real-world temporal processing tasks from a diverse range of application scenarios, each capturing rich temporal dynamics across multiple timescales. Utilizing NSA, we conduct extensive comparisons of recently introduced spiking neuron models and neural architectures, presenting comprehensive baselines in terms of task performance, training speed, memory usage, and energy efficiency. Our findings emphasize an urgent need for efficient SNN designs that can consistently deliver high performance across tasks with varying temporal complexities while maintaining low computational costs. NSA enables systematic tracking of advancements in neuromorphic algorithm research and paves the way for developing effective and efficient neuromorphic temporal processing systems.

1 Introduction

Temporal processing is fundamental for intelligent systems to interpret time-varying sensory signals, facilitating accurate and timely decision-making in dynamic environments. Neuromorphic computing holds immense potential for processing these signals with ultra-low energy consumption and low latency. Spiking Neural Networks (SNNs), inspired by the computational principles of biological brains [Maass, 1997], serve as a cornerstone of neuromorphic computing. While SNNs have achieved performance on par with traditional Artificial Neural Networks (ANNs) in many static image classification tasks, demonstrating significantly improved energy efficiency and reduced latency [Davies *et al.*, 2018; Pei *et al.*, 2019; Ma *et al.*, 2024; Yang *et al.*, 2024], their capacities in processing temporal signals remain inferior to those of ANNs. Recently, numerous SNN approaches have been proposed, making substantial progress in bridging this gap. These advancements include enriching neural dynamic heterogeneity [Yin *et al.*, 2021; Zheng *et al.*, 2024], increasing neuronal structural complexity [Zhang *et al.*, 2024; He *et al.*, 2024; Sun *et al.*, 2024], and developing temporal parallelization methods for improved training efficiency [Fang *et al.*, 2023; Chen *et al.*, 2024].

Despite this progress, advancements in this research area are hindered by the absence of effective benchmarks. Existing SNN benchmarks predominantly focus on visual classification [LeCun *et al.*, 1998; Li *et al.*, 2017; Amir *et al.*, 2017] and keyword spotting [Warden, 2018; Cramer *et al.*, 2020] tasks. While these benchmarks have been valuable over the past decade in advancing neuromorphic computing research, they play a limited role in fostering developments in neuromorphic temporal processing due to three primary limitations. First, the model performance on the current SNN benchmarks is largely saturated, and these benchmarks fail to capture the rich temporal dynamics inherent in real-world signals, which typically encompass multiple timescales. Second, existing benchmarks do not adequately represent the diverse range of temporal processing scenarios that closely align with the interests and objectives of neuromorphic computing research. Third, inconsistent comparisons across studies and the absence of evaluations on critical efficiency metrics result in biased assessments of model practicality.

To address these limitations, we propose a comprehensive benchmark called **Neuromorphic Sequential Arena (NSA)**, designed to establish an effective, versatile, and application-oriented evaluation framework for neuromorphic temporal processing. Firstly, NSA encompasses a broad range of tasks that reflect real-world temporal processing scenarios across application areas of significant interest to the neuromorphic research community, including human-computer interaction, speech processing, robotics, and biomedical appli-

*Corresponding author: Jibin Wu.

cations. Furthermore, to ensure that these tasks possess an adequate level of temporal complexity suitable for SNNs, we introduce a novel tool for analyzing the temporal dependencies required to address a given task, referred to as the Segregated Temporal Probe (STP). This tool has been employed to validate the effectiveness of both commonly used neuromorphic datasets and the proposed NSA in benchmarking the temporal processing capacity of SNNs. Finally, using NSA, we conduct a comprehensive comparative study of recently introduced spiking neuron models and neural architectures in terms of task performance, training speed, memory usage, and energy efficiency. By providing a side-by-side comparison of these baselines, NSA serves as a valuable foundation for understanding the current status and practicality of existing methods. Collectively, NSA is designed as an evolving framework capable of integrating emerging neuromorphic algorithms, thereby fostering advancements toward more effective and efficient neuromorphic temporal processing systems.

2 Neuromorphic Sequential Arena (NSA)

The NSA comprises seven tasks that span a diverse array of real-world scenarios relevant to neuromorphic research. These tasks, characterized by varying levels of temporal complexity, serve as an effective benchmark for assessing the temporal processing capacities of different SNN approaches. In the following, we detail our design principles, task formulation, and STP tool used to assess task effectiveness.

2.1 Design Principles

This section outlines the five foundational principles for the NSA design:

- **Neuromorphic relevance.** Tasks should highlight the advantages of neuromorphic solutions, such as energy efficiency, low latency, and robustness.
- **Temporal complexity.** Tasks should reflect rich temporal dynamics inherent in real-world scenarios, requiring models to establish temporal dependencies across multiple timescales.
- **Challenging.** Tasks should exhibit an adequate level of difficulty, highlighting distinguishable performance among existing SNN approaches while offering significant opportunities for improvement.
- **Training resource.** Tasks should account for training time and GPU resource constraints to ensure accessibility for both researchers and practitioners. In particular, for SNN approaches that rely on temporally serial simulation, it is crucial that the training can be completed within a reasonable amount of training time.
- **Application versatility.** Tasks should encompass a wide range of real-world application scenarios with distinct requirements and data characteristics, bridging the gap between theoretical advancements and practical applications.

2.2 Tasks

In the following, we will introduce the suite of tasks included in the proposed NSA benchmark, meticulously designed following the principles outlined above. We provide a comprehensive overview of these tasks, including detailed task descriptions, application scenarios, and specific capacities they evaluate. The task characteristics, evaluation metrics, and dataset configurations are summarized in Table 1. Detailed preprocessing techniques for these tasks are provided in the Supplementary Materials. To facilitate benchmarking efforts, we provide a comprehensive **open-source library**[*] that allows seamless integration of novel spiking neuron models and neural architectures, and ensures consistent evaluations across different methods.

Autonomous Localization (AL)

Due to the inherent characteristics of low latency and high energy efficiency, neuromorphic systems present considerable potential for robotic control. A fundamental challenge in this domain is predicting the current state of the robot after executing a sequence of actions, which is crucial for overcoming sensory feedback delays and promptly adapting to external perturbations. Addressing this challenge requires a neuromorphic system to establish long-term temporal dependencies, effectively modeling the intricate relationship between past actions and the present state.

Given the above requirement and the significance of this task, we propose a new synthetic dataset called AL. AL simulates a scenario in which a mobile robot must determine whether it is positioned on the left or right plane following a sequence of ordered actions, including 'turn left', 'turn right', 'go straight', and 'stop'. This constitutes a binary classification task with customized sequence lengths and action distributions, facilitating an effective evaluation of the capacity of SNNs to establish temporal dependencies across various timescales.

Human Activities Recognition (HAR)

The HAR task focuses on identifying human activity patterns from time-series data collected by wearable sensors. This task represents a significant application of neuromorphic computing, encompassing real-time robotic locomotion control, pose estimation, and surveillance, where accurate and efficient temporal processing is essential. HAR requires the modeling of fine-grained temporal trajectories of human activities embedded within noisy sensor signals. Therefore, HAR is particularly well-suited for evaluating the capacity of SNNs to capture complex short-term temporal dependencies and to demonstrate their robustness against unpredictable noise conditions.

In this task, we utilize data samples from the WISDM dataset [Weiss, 2019], which includes gyroscope data collected from a smartwatch. The data is recorded at 20 Hz and segmented into 10-second intervals, resulting in a sequence length of 200. This task is intentionally challenging, requiring models to accurately classify 18 actions that may exhibit subtle differences, such as 'eating chips' or 'eating pizza'.

Electroencephalogram Motor Imagery (EEG-MI)

We further evaluate the effectiveness of SNNs in brain-computer interfaces and real-time cognitive monitoring

[*]The source code and supplementary materials are publicly available at https://github.com/liyc5929/neuroseqbench.

Task	Dataset	Sequence length	Metric	Backbone model	Training samples	Testing samples
AL	AL	User-defined	Acc.	MLP	50,000	5,000
HAR	WISDM	200	Acc.	MLP	26,048	6,512
EEG-MI	OpenBMI	500	Acc.	MLP	17,280	4,320
SSL	SLoClas	500	Acc.	MLP	37,426	8,969
ALR	DVS-Lip	200	Acc.	MLP	14,896	4,975
AD	N-DNS	751/3,751	SI-SNR	Spiking-FullSubNet	60,000	341
ASR	AISHELL	76–505	CER	VGG-MLP	360,294	7,176

Table 1: Summary of tasks characteristics, evaluation metrics, and other configurations in the proposed NSA benchmark.

through the EEG-MI task. This task focuses on decoding motion imagery from EEG sequences, which is particularly challenging due to the sparse, noisy, and highly dynamic nature of EEG signals. Therefore, it serves as an effective benchmark for assessing the capacity of SNNs to capture both temporal and spatial dependencies with a high degree of robustness.

Specifically, we utilize motion imagery data from the OpenBMI dataset [Lee *et al.*, 2019], which comprises 62-channel EEG recordings collected from 52 subjects at a sampling rate of 1 kHz. The model is tasked with performing a binary classification aimed at distinguishing between imagery trials of left- and right-hand grasping. To ensure manageable training costs, we downsample the sequence length to 500. To mitigate the risk of overfitting due to subject-specific bias, we employ a cross-trail validation approach, in which all samples are randomly partitioned into training and testing sets.

Sound Source Localization (SSL)

Identifying the origin of a sound source within a noisy environment is a crucial survival skill for animals and holds significant importance for applications such as hearing aids and robotics. In this task, we aim to evaluate the temporal processing capacity and noise robustness of SNN approaches. To this end, we adopt the SLoClas dataset [Qian *et al.*, 2021], which contains 4-channel audio recordings from a single sound source positioned at azimuth angles ranging from 0° to 360° in 5° increments, resulting in a 72-class classification problem. Additionally, directional background noise fragments are introduced to the raw audio signals at a challenging signal-to-noise (SNR) ratio of 0. The resulting audio signals are segmented into sequences of length 500, then fed directly into the SNNs without any spectral preprocessing. This task poses a specific challenge for temporal processing, as the model must learn to map the temporal delays between different audio channels to their corresponding azimuth angles in the presence of background noise.

Automatic Lip-Reading (ALR)

ALR aims to recognize spoken words from a speaker's lip movements and plays a pivotal role in various real-world applications, such as video surveillance and speech recognition in noisy environments. Dynamic Vision Sensor (DVS) cameras, characterized by high dynamic range and low latency, have emerged as ideal visual front-ends for capturing the fine-grained lip movements essential for the ALR task. Their event-based outputs are particularly well-suited for neuromorphic systems.

In this task, we adopt the DVS-Lip dataset [Tan *et al.*, 2022], which consists of 100 spoken word classes captured by the DAVIS346 event camera from 40 individuals. The task is particularly challenging, as the training and testing sets include different individuals, requiring the SNN model to generalize across unseen speakers. Moreover, the output is decoded solely from the spiking activities generated at the final time step, requiring models to extract and retain crucial spatiotemporal features over an extended period. To ensure manageable training costs, each sample is segmented into 200 temporal bins and center-cropped to a resolution of 88×88 pixels.

Audio Denoising (AD)

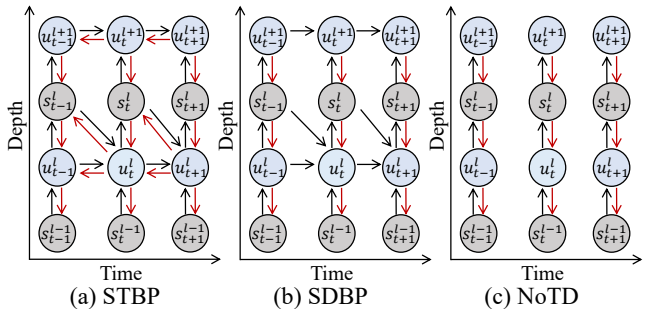
Removing noise from received audio signals to enhance overall quality is critical for many edge applications, such as hands-free communication and hearing aids. Given the low-power and real-time processing requirements of these tasks, neuromorphic solutions present a highly promising approach. Notably, among all tasks within the proposed NSA benchmark, AD stands out as a regression problem, providing a unique opportunity to evaluate the expressive power of SNN models in capturing subtle and continuous temporal variations.

In this task, we utilize the publicly available Intel Neuromorphic Deep Noise Suppression (N-DNS) Challenge dataset [Timcheck *et al.*, 2023]. This dataset comprises clean speech samples in multiple languages (i.e., English, German, French, Spanish, and Russian) as well as noise samples collected from diverse acoustic environments. We employ the official synthesizer script to generate a 495-hour training subset and a 5-hour validation subset. All audio samples are synthesized at a 16 kHz sampling rate and segmented into a uniform length of 30 seconds. We perform audio preprocessing steps in accordance with the winning entry of the N-DNS Challenge [Hao *et al.*, 2024] and adopt their model architecture as the default. For model evaluation, we adopt the Scale-Invariant Signal-to-Noise Ratio (SI-SNR) metric.

Automatic Speech Recognition (ASR)

ASR transcribes spoken language from audio into text, serving as a foundational technology for various applications, including voice assistants, transcription services, and speech translation tools. This task presents significant challenges for SNNs due to varying sequence lengths, speaker characteristics, and acoustic conditions.

In this task, we utilize the AISHELL dataset [Bu *et al.*, 2017], an open-sourced Mandarin speech corpus comprising approximately 170 hours of speech data from 400 speakers, encompassing a wide range of accents and speaking styles. Its moderate data scale and high speaker diversity render it



(a) STBP (b) SDBP (c) NoTD

Figure 1: Comparison of the three training algorithms in STP.

an effective testbed for assessing the model's capacity to handle linguistic and acoustic variations. Model performance is evaluated using the Character Error Rate (CER).

2.3 Segregated Temporal Probe (STP)

To elucidate the limited effectiveness of commonly used neuromorphic benchmarks in evaluating the temporal processing capacity of SNNs and to demonstrate the efficacy of NSA, we introduce STP. STP assesses the contributions of establishing temporal dependencies by quantifying the impact of isolating forward and backward temporal processing pathways of spiking neurons on task performance. As illustrated in Figure 1, STP consists of three evaluation modules, each corresponding to a specific training algorithm: Spatio-Temporal Backpropagation (STBP) [Wu *et al.*, 2018], Spatial Domain Backpropagation (SDBP), and No Temporal Domain (NoTD). Details of these three algorithms are outlined below, using the Leaky Integrate-and-Fire (LIF) neuron model [Burkitt, 2006] as an example.

The LIF neuron accumulates input spikes $\boldsymbol{s}^{l-1}[t]$ from the preceding layer $l-1$ into its membrane potential $\boldsymbol{u}^{l}[t]$. When $\boldsymbol{u}^{l}[t]$ surpasses a threshold V_{th}, it triggers a spike and subsequently resets to the resting potential. The dynamics of a LIF neuron can be formulated as:

$$\boldsymbol{u}^{l}[t] = \underbrace{\lambda \boldsymbol{u}^{l}[t-1](1-\boldsymbol{s}^{l}[t-1])}_{\text{Forward temporal propagation}} + \boldsymbol{W}^{l}\boldsymbol{s}^{l-1}[t], \quad (1)$$

$$\boldsymbol{s}^{l}[t] = \Theta(\boldsymbol{u}^{l}[t] - V_{\text{th}}), \quad (2)$$

where λ determines the decay rate of $\boldsymbol{u}^{l}[t]$ over time, $\boldsymbol{W}^{l}$ is the synaptic weight matrix, and $\Theta(\cdot)$ is the Step function. The recursive update in Eq. (1) allows the LIF neuron to iteratively propagate temporal information from $\boldsymbol{u}^{l}[t-1]$ to $\boldsymbol{u}^{l}[t]$, enabling the integration of temporal information over time.

STBP preserves this temporal propagation in both forward and backward passes, where the gradient of the loss with respect to weights is computed as:

$$\frac{\partial \mathcal{L}}{\partial \boldsymbol{W}^{l}} = \sum_{t=1}^{T} \frac{\partial \mathcal{L}}{\partial \boldsymbol{u}^{l}[t]} \frac{\partial \boldsymbol{u}^{l}[t]}{\partial \boldsymbol{W}^{l}} = \sum_{t=1}^{T} \boldsymbol{\delta}^{l}[t]^{\top} \boldsymbol{s}^{l-1}[t]^{\top}, \quad (3)$$

$$\boldsymbol{\delta}^{l}[t] = \underbrace{\boldsymbol{\delta}^{l}[t+1]\frac{\partial \boldsymbol{u}^{l}[t+1]}{\partial \boldsymbol{u}^{l}[t]}}_{\text{Backward temporal propagation}} + \boldsymbol{\delta}^{l+1}[t]\frac{\partial \boldsymbol{u}^{l+1}[t]}{\partial \boldsymbol{u}^{l}[t]}, \quad (4)$$

where $\boldsymbol{\delta}^{l}[t] \triangleq \frac{\partial \mathcal{L}}{\partial \boldsymbol{u}^{l}[t]}$, which is the backpropagated gradient error along both spatial and temporal dimensions.

SDBP only retains the temporal propagation in the forward pass but omits its impact on the backward pass. This prevents the gradient from propagating across time, thereby limiting the learning of temporal dependency. The modified gradient error is:

$$\hat{\boldsymbol{\delta}}^{l}[t] = \begin{cases} \frac{\partial \mathcal{L}}{\partial \boldsymbol{u}^{L}[t]}, & l = L, \\ \hat{\boldsymbol{\delta}}^{l+1}[t]\frac{\partial \boldsymbol{u}^{l+1}[t]}{\partial \boldsymbol{u}^{l}[t]}, & l < L. \end{cases} \quad (5)$$

NoTD removes temporal propagation in both forward and backward passes, processing each time step independently. The resulting neuronal dynamic is simplified to:

$$\boldsymbol{u}^{l}[t] = \boldsymbol{W}^{l}\boldsymbol{s}^{l-1}[t], \quad (6)$$

which excludes the update of $\boldsymbol{u}^{l}[t]$ compared to Eq. (1), rendering it unable to integrate any temporal information across time. The gradient in NoTD is the same as Eq. (5) in SDBP.

By training an SNN using the three algorithms described and comparing their performance gaps, we can quantify the effectiveness of a specific task in assessing the temporal processing capacity of the SNN:

1. A similar performance between STBP and SDBP indicates that the task fails to leverage temporal propagation in the backward pass to establish meaningful temporal dependencies, and, therefore, is ineffective in evaluating temporal processing capacity.
2. A comparable performance between NoTD and STBP implies that the task can be solved without leveraging any temporal information, making it unsuitable as a benchmark for temporal processing.
3. In contrast, significant performance degradation in SDBP and NoTD compared to STBP indicates that the task contains rich temporal information and can effectively evaluate temporal processing capacity.

3 Results

In this section, we first employ STP to validate the effectiveness of the existing and the proposed NSA benchmarks in evaluating the temporal processing capacity of SNNs. Following this, we conduct a comprehensive comparison of recently proposed spiking neuron models and neural architectures using NSA, thereby illuminating the current status of SNNs in temporal processing. Additionally, we incorporate deployment-critical efficiency metrics to further benchmark these models, including training speed, memory usage, and energy consumption, to assess their practicality for real-world applications. For all Tables, **bold values** indicate the best performance, and underlined values represent the second best. Details on experimental configurations and hyperparameters are provided in Supplementary Materials.

3.1 Benchmark Effectiveness Validation

We first apply the proposed STP tool to evaluate 12 commonly used benchmarks in the neuromorphic community (see

Method	AL Acc. (%) ↑	HAR Acc. (%) ↑	EEG-MI Acc. (%) ↑	SSL Acc. (%) ↑	ALR Acc. (%) ↑	AD SI-SNR (dB) ↑	ASR CER (%) ↓
STBP	**63.52**	**81.27**	**65.20**	**8.88**	**17.83**	**11.47**	**20.70**
SDBP	58.52 (-5.00)	76.29 (-4.98)	57.48 (-7.72)	5.79 (-3.09)	2.47 (-15.36)	10.18 (-1.29)	26.30 (+5.60)
NoTD	53.34 (-10.18)	67.75 (-13.52)	52.48 (-12.72)	3.45 (-5.43)	2.01 (-15.82)	9.47 (-2.00)	27.10 (+6.40)

Table 2: Validating the effectiveness of NSA in evaluating the temporal processing capacity of SNNs.

Architecture	Neuron model	AL Acc. (%) ↑	HAR Acc. (%) ↑	EEG-MI Acc. (%) ↑	SSL Acc. (%) ↑	ALR Acc. (%) ↑	AD SI-SNR (dB) ↑	ASR CER (%) ↓	Average rank ↓
SFNN	LIF	54.40	81.27	65.20	8.88	17.83	11.47	20.70	6.6
	CE-LIF	59.58	79.71	70.56	12.21	48.32	-*	17.70	5.3
	LTC	77.40	80.97	64.42	10.33	48.93	**14.36**	15.90	4.3
	SPSN	72.02	86.73	**76.46**	20.30	45.73	13.00	18.50	3.7
	PMSN	**87.42**	**88.31**	75.00	**75.35**	**57.43**	14.35	17.70	**1.9**
SRNN	LIF	56.50	77.32	64.68	6.17	34.71	9.36	15.70	6.4
	CE-LIF	60.62	80.83	74.35	12.72	51.64	-*	**15.30**	3.7
	LTC	79.04	82.03	69.16	16.37	56.64	14.29	16.30	3.1

* These models are not applicable due to the inconsistent sequence lengths between the training and test phases.

Table 3: Results of NSA benchmark for different spiking neuron models with SFNN or SRNN architectures.

Architecture	AL Acc. (%) ↑	HAR Acc. (%) ↑	EEG-MI Acc. (%) ↑	SSL Acc. (%) ↑	ALR Acc. (%) ↑	AD SI-SNR (dB) ↑	ASR CER (%) ↓	Average rank ↓
SFNN	54.40	81.27	65.20	8.88	17.83	11.47	20.70	5.9
SRNN	56.50	77.32	64.68	6.17	34.71	9.36	15.70	5.9
GSN	67.22	82.31	68.85	10.34	21.17	**14.45**	**14.30**	3.6
Spiking TCN	69.88	82.41	75.58	40.60	**47.14**	12.77	17.10	3.1
Spike-Driven Transformer	58.00	71.07	69.12	5.75	39.62	9.86	36.80	5.7
Binary S4D	**81.44**	**89.37**	**78.61**	79.34	44.80	14.21	15.20	**1.7**
GSU	80.42	89.16	77.43	**82.39**	41.35	14.05	14.40	2.1

Table 4: Results of NSA benchmark for different neural architectures using LIF neurons.

Supplementary Materials for details). Our results indicate comparable performance between STBP and SDBP, suggesting that their effectiveness in assessing the temporal processing capacities of SNNs is limited. Subsequently, we perform the same study on the proposed NSA benchmark to validate its effectiveness. As shown in Table 2, both SDBP and NoTD exhibit substantial performance degradation compared to STBP across all tasks in NSA. This suggests that the seven selected tasks contain essential temporal dependencies that must be effectively captured to attain high performance. Consequently, the proposed NSA serves as a more effective benchmark for neuromorphic temporal processing.

3.2 Performance Benchmarking of SNN Models

To elucidate the current status of SNN models in temporal processing, we further evaluate five spiking neuron models notable for their enhanced temporal processing capacities: LIF, Context Embedding LIF (CE-LIF) [Chen *et al.*, 2023], Liquid Time-Constant (LTC) [Yin *et al.*, 2023], Sliding Parallel Spiking Neuron (SPSN) [Fang *et al.*, 2023], and Parallel Multicompartment Spiking Neuron (PMSN) [Chen *et al.*, 2024]. In addition to the commonly used feedforward (SFNN) and recurrent (SRNN) networks [Bellec *et al.*, 2018], we also benchmark five advanced neural architectures that excel in temporal processing, including Gated Spiking Neuron (GSN) [Hao *et al.*, 2024], Spiking Temporal Convolution Network (Spiking TCN) [Bai *et al.*, 2018], Spike-Driven Transformer [Yao *et al.*, 2024], Binary S4D [Stan and Rhodes, 2024], and Gated Spiking Unit (GSU) [Stan and Rhodes, 2024]. To ensure a fair comparison, all models are configured with a comparable number of trainable parameters for each task. For architectures like Spiking TCN and Spike-Driven Transformer, which are originally designed to repeat each time step D times to enhance their model representation power, we set $D = 1$ to facilitate fair comparisons with other models.

Our experimental results demonstrate that the performance of spiking neuron models varies significantly across tasks. LTC shows only marginal improvements in noise-intensive tasks such as HAR, EEG-MI, and SSL, attributable to its input-dependent mechanism, which is highly susceptible to noise accumulation over time. In contrast, PMSN and SPSN neurons exhibit stronger noise robustness in these tasks. Additionally, PMSN and LTC excel in AL and ALR tasks requiring long-term memory, showcasing superior temporal retention capacities over extended periods. Furthermore, LTC and PMSN demonstrate high expressiveness in the AD task, while CE-LIF with recurrent connections stands out in the

ASR task, exhibiting remarkable generalizability. Overall, PMSN emerges as the top-performing spiking neuron model on NSA, striking a balance between robustness, temporal dependency establishment, and generalizability across tasks.

For neural architectures, Binary S4D and its variant GSU exhibit strong noise robustness in HAR, EEG-MI, and SSL tasks, outperforming other architectures in handling noisy inputs. In contrast, the GSN model achieves notable gains in AD and ASR tasks, which require high model expressiveness and generalizability. This result suggests the high effectiveness of GSN in capturing diverse, fine-grained temporal patterns. Conversely, SRNNs and Spike-Driven Transformer fall short in performance compared to other models. We noticed that SRNNs often suffer from training instability and convergence issues in our experiments, resulting in comparable or worse performance than SFNNs. The Spike-Driven Transformer struggles to establish temporal dependencies effectively due to its binary activations (i.e., $D = 1$), which significantly constrains its expressiveness. Meanwhile, this architecture is typically designed to work with large datasets and networks, the moderate dataset size of NSA and the small parameter count in our evaluation may constrain the full exploitation of its architectural advantages.

3.3 Efficiency Benchmarking of SNN Models

We further report the efficiency metrics of the evaluated SNN models on the AL task, including training speed, GPU memory usage, and energy efficiency, which are critical factors for practical implementation. To ensure a fair comparison, all evaluations are conducted using a batch size of 256 and all evaluated SNN models are configured with a uniform network dimension, comprising two hidden layers with 256 channels each. Training speed and memory costs are evaluated with sequence lengths of $\{200, 400, 800\}$, while energy efficiency is assessed with a sequence length of 400. The comparative results are summarized in Table 5.

Training Speed

Our evaluation results on NSA exhibit notable training speed differences between serial and parallel models. Serial models, including LIF, CE-LIF, LTC, and GSN, process temporal data sequentially, maintaining consistent but inherently slow training speeds per time step, regardless of sequence lengths. Consequently, their total training time scales linearly with the sequence length, leading to high computational costs for processing long sequences. This limitation becomes even more pronounced when involving complex neuronal dynamics. For instance, LIF achieves the fastest training speed among serial models, whereas LTC achieves the slowest, highlighting that inefficient training in serial models poses a significant bottleneck for long sequence processing.

Parallel models, on the contrary, demonstrate high computational efficiency by processing data in multiple time steps simultaneously. Our results show that recently proposed parallel models, including SPSN, PMSN, Spiking TCN, binary S4D, and GSU, achieve approximately $3\times$ speedup compared to serial models. However, the Spike-Driven Transformer fails to achieve noticeable speedups over the LIF model due to its high computational complexity. Additionally, the training speed of parallel models varies with sequence length, reflecting differences in their underlying parallelization strategies. Specifically, SPSN exhibits a decline in training speed as the sequence length increases, resulting in an exponentially growing total training time. This slowdown arises from the design of its receptive kernel, which expands proportionally with sequence length to capture long-range temporal dependencies. Consequently, the neuron dynamics incur a quadratic time complexity of $\mathcal{O}(L^2)$. In contrast, other parallel models sustain consistent training speeds across varying sequence lengths, owing to their more efficient linear-time complexity of $\mathcal{O}(L)$.

Memory Usage

Despite enhanced temporal processing capacity offered by many recent SNN models, our finding reveals that these advancements come with a significant increase in memory consumption. Specifically, while parallel computing models accelerate training speeds, they typically incur substantially higher memory usage, trading off between space and time. This inefficiency arises from their need to store additional states to capture temporal dependencies in parallel. Among these parallel models, the Spike-Driven Transformer stands out as the most memory-intensive due to the expanded embedding space involved in the self-attention mechanism. In contrast, PMSN, S4D, and GSU consume comparatively less memory, benefiting from their compact representation of model states. However, they still exceed the memory requirements of serial approaches due to the additional storage needed for their parallelized operations. SPSN employs a 1-D temporal convolution kernel to capture local temporal features without the need to buffer extra transient states, making it a highly memory-efficient parallel architecture comparable to the LIF model. In contrast, serial models generally consume less memory as they store fewer transient states across time. The only exception is LTC, whose poor memory efficiency stems from its complex computation graph involved in gating computations. Notably, memory consumption for all evaluated models scales linearly with sequence length, regardless of structure or computational strategy. It highlights SNNs' memory efficiency in handling extended temporal sequences, as each time step contributes a fixed and predictable amount of memory without exponential growth.

Energy Efficiency

The high energy efficiency of SNNs promises to enable efficient temporal processing at the edge. To present a comprehensive assessment of the energy efficiency of advanced SNN models, we conduct a quantitative analysis of the number of Multiply-Accumulate (MAC) and Accumulate (AC) operations per inference time step and per sample. Additionally, we derive the average empirical energy consumption for each model based on the experimental data. Detailed calculations of energy cost can be found in the Supplementary Materials.

The results presented in Table 5 suggest that most advanced SNN models improve temporal processing capacity at the cost of increased energy consumption. Among the evaluated models, CE-LIF, PMSN, and GSU stand out for achieving a favorable trade-off between performance and energy efficiency, making them more suitable for deployment in

Architecture	Neuron model	Training speed (k step/s) ↑			Memory consumption (GB) ↓			Energy efficiency		
		200	400	800	200	400	800	**ACs** (k) ↓	**MACs** (k) ↓	**Empirical cost** (nJ) ↓
SFNN	LIF	1.91 (1.0)	1.91 (1.0)	2.04 (1.0)	**0.49 (1.0)**	**0.98 (1.0)**	**1.96 (1.0)**	3.72 (1.0)	0.26 (1.0)	**4.52 (1.0)**
	CE-LIF	1.25 (0.7)	1.31 (0.7)	1.37 (0.7)	0.59 (1.2)	1.18 (1.2)	2.35 (1.2)	3.22 (0.9)	0.77 (3.0)	6.43 (1.4)
	LTC	0.63 (0.3)	0.66 (0.3)	0.68 (0.3)	1.43 (2.9)	2.84 (2.9)	5.68 (2.9)	**0.93 (0.3)**	262.91 (1,011)	1,210.23 (268)
	SPSN	5.04 (2.6)	4.14 (2.2)	3.21 (1.6)	0.50 (1.0)	1.01 (1.0)	2.01 (1.0)	1.56 (0.4)	32.77 (126)	152.14 (33)
	PMSN	6.34 (3.3)	6.31 (3.3)	6.24 (3.1)	1.23 (2.5)	2.46 (2.5)	4.92 (2.5)	10.27 (2.8)	4.66 (18)	30.68 (6.8)
SRNN	LIF	1.19 (0.6)	1.25 (0.7)	1.30 (0.6)	0.54 (1.1)	1.08 (1.1)	2.16 (1.1)	22.91 (6.2)	0.26 (1.0)	21.79 (4.8)
	CE-LIF	0.92 (0.5)	0.98 (0.5)	1.00 (0.5)	0.64 (1.3)	1.28 (1.3)	2.55 (1.3)	8.26 (2.2)	0.77 (3.0)	10.96 (2.4)
	LTC	0.55 (0.3)	0.57 (0.3)	0.57 (0.3)	1.43 (2.9)	2.84 (2.9)	5.68 (2.9)	2.12 (0.6)	262.91 (1,011)	1,211.30 (268)
GSN	LIF	0.88 (0.5)	0.92 (0.5)	0.92 (0.5)	0.64 (1.3)	1.28 (1.3)	2.55 (1.3)	27.33 (7.3)	1.28 (4.9)	30.48 (6.7)
Spiking TCN		4.92 (2.6)	4.99 (2.6)	5.01 (2.5)	0.66 (1.4)	1.30 (1.3)	2.58 (1.3)	50.49 (11)	**0.00 (0.0)**	56.10 (15)
Spike-Driven Transformer		1.86 (1.0)	1.99 (1.0)	2.06 (1.0)	3.07 (6.3)	6.10 (6.2)	12.16 (6.2)	214.70 (58)	**0.00 (0.0)**	193.22 (42)
Binary S4D		**7.04 (3.7)**	**7.07 (3.7)**	**7.33 (3.6)**	1.26 (2.6)	2.51 (2.6)	5.00 (2.6)	44.49 (12)	5.43 (21)	65.01 (14)
GSU		5.93 (3.1)	5.91 (3.1)	5.92 (2.9)	1.48 (3.0)	2.95 (3.0)	5.88 (3.0)	6.37 (1.7)	4.92 (19)	28.36 (6.3)

Table 5: Comparison of SNN models in terms of the training speed, memory consumption, and energy efficiency. Values in brackets indicate ratios relative to the baseline LIF-SFNN model.

resource-constrained edge systems. It is noteworthy that LTC presents two orders of magnitude higher energy consumption than standard LIF models, which can be attributed to the hundreds of thousands of MAC operations required for computing its temporal dynamics. Such high energy costs comparable to traditional ANN methods severely limit its practicality for energy-constrained systems. These observations highlight the critical role of efficiency evaluation in model design. To unleash the full potential of neuromorphic systems in temporal processing, we argue that the development of SNN models must prioritize not only the enhancement of task accuracy but also the maintenance of ultra-low energy consumption, thereby ensuring their competitiveness in real-world applications.

4 Related Works: Current SNN Benchmarking Practices

Existing benchmarks commonly used for evaluating SNNs can be divided into four categories, each with its own limitations that hinder their ability to effectively assess the temporal processing capacity of SNNs. The first category includes static image recognition tasks [LeCun *et al.*, 1998; Krizhevsky and Hinton, 2009], where identical images are repeated along the time axis, lacking any meaningful temporal dynamics. The second category comprises event-based visual classification tasks recorded by DVS cameras [Li *et al.*, 2017; Amir *et al.*, 2017; Zhou *et al.*, 2024; Wang *et al.*, 2022; Wang *et al.*, 2024]. While these datasets impose artificial saccadic motion on static images or capture simple moving objects, their limited temporal dynamics result in performance that primarily emphasizes spatial pattern recognition rather than the establishment of long-range temporal dependencies. The third category involves keyword spotting tasks, which encompass both frame-based [Warden, 2018] and spike-based [Cramer *et al.*, 2020] audio inputs. While these datasets contain richer temporal dynamics, effective decisions can often be made by integrating only short-term temporal features, making these datasets insufficiently challenging to evaluate the temporal processing capacity of SNNs. More recently, several preliminary efforts have been made to apply SNNs to long-term language modeling tasks [Tay *et al.*, 2020]. Despite the complex temporal dependencies inherent in these tasks, the high training costs associated with these models render them unsuitable for evaluating many existing SNN approaches. Furthermore, addressing such tasks typically necessitates models with a substantial number of parameters, which are not feasible for deployment on current neuromorphic hardware, thereby limiting their utility as benchmarks for SNNs at this stage of development.

5 Discussion and Conclusion

In this work, we present NSA, an effective, versatile, and application-oriented benchmark designed to comprehensively evaluate the temporal processing capacities of SNNs across diverse application scenarios. To ensure rigorous and reliable assessment, we integrate a temporal dependency analysis tool, STP, into NSA to quantify the effectiveness of the benchmark. Our comparative analysis underscores both the progress and the challenges in neuromorphic temporal processing. While advanced spiking neuron models and neural architectures demonstrate remarkable improvements in task performance, many of them struggle to maintain efficiency in training speed, memory usage, and energy consumption, which are crucial constraints for real-world applications. Our findings underscore the urgent need to develop effective SNN models capable of robustly processing temporal data while maintaining high energy efficiency. While this paper provides a limited evaluation of SNN approaches due to time and resource constraints, we encourage the community to expand the scope of evaluations using NSA. We envision NSA as an effective and adaptive temporal benchmarking framework capable of addressing the evolving needs of the community. We hope this benchmark will inspire further advancements in neuromorphic temporal processing research, thereby paving the way for more capable, robust, and efficient neuromorphic solutions for real-world applications.

Acknowledgments

This work was partially supported by the National Natural Science Foundation of China (Grant No. 62306259 and U21A20512), the Research Grants Council of the Hong Kong SAR (Grant No. PolyU25216423, PolyU11211521, PolyU15218622, PolyU15215623, and C5052-23G), and The Hong Kong Polytechnic University (Project IDs: P0043563, P0046094).

Contribution Statement

X. Chen and C. Ma contributed equally to this work.

References

[Amir *et al.*, 2017] Arnon Amir, Brian Taba, David Berg, Timothy Melano, Jeffrey McKinstry, Carmelo Di Nolfo, Tapan Nayak, Alexander Andreopoulos, Guillaume Garreau, Marcela Mendoza, et al. A low power, fully event-based gesture recognition system. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 7243–7252, 2017.

[Bai *et al.*, 2018] Shaojie Bai, J Zico Kolter, and Vladlen Koltun. An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. *arXiv preprint arXiv:1803.01271*, 2018.

[Bellec *et al.*, 2018] Guillaume Bellec, Darjan Salaj, Anand Subramoney, Robert Legenstein, and Wolfgang Maass. Long short-term memory and learning-to-learn in networks of spiking neurons. *Advances in neural information processing systems*, 31, 2018.

[Bu *et al.*, 2017] Hui Bu, Jiayu Du, Xingyu Na, Bengu Wu, and Hao Zheng. Aishell-1: An open-source mandarin speech corpus and a speech recognition baseline. In *2017 20th conference of the oriental chapter of the international coordinating committee on speech databases and speech I/O systems and assessment (O-COCOSDA)*, pages 1–5. IEEE, 2017.

[Burkitt, 2006] Anthony N Burkitt. A review of the integrate-and-fire neuron model: I. homogeneous synaptic input. *Biological cybernetics*, 95:1–19, 2006.

[Chen *et al.*, 2023] Xinyi Chen, Jibin Wu, Huajin Tang, Qinyuan Ren, and Kay Chen Tan. Unleashing the potential of spiking neural networks for sequential modeling with contextual embedding. *arXiv preprint arXiv:2308.15150*, 2023.

[Chen *et al.*, 2024] Xinyi Chen, Jibin Wu, Chenxiang Ma, Yinsong Yan, Yujie Wu, and Kay Chen Tan. Pmsn: A parallel multi-compartment spiking neuron for multi-scale temporal processing. *arXiv preprint arXiv:2408.14917*, 2024.

[Cramer *et al.*, 2020] Benjamin Cramer, Yannik Stradmann, Johannes Schemmel, and Friedemann Zenke. The heidelberg spiking data sets for the systematic evaluation of spiking neural networks. *IEEE Transactions on Neural Networks and Learning Systems*, 33(7):2744–2757, 2020.

[Davies *et al.*, 2018] Mike Davies, Narayan Srinivasa, Tsung-Han Lin, Gautham Chinya, Yongqiang Cao, Sri Harsha Choday, Georgios Dimou, Prasad Joshi, Nabil Imam, Shweta Jain, et al. Loihi: A neuromorphic manycore processor with on-chip learning. *IEEE Micro*, 38(1):82–99, 2018.

[Fang *et al.*, 2023] Wei Fang, Zhaofei Yu, Zhaokun Zhou, Ding Chen, Yanqi Chen, Zhengyu Ma, Timothée Masquelier, and Yonghong Tian. Parallel spiking neurons with high efficiency and ability to learn long-term dependencies. In *Advances in Neural Information Processing Systems*, volume 36, pages 53674–53687, 2023.

[Hao *et al.*, 2024] Xiang Hao, Chenxiang Ma, Qu Yang, Jibin Wu, and Kay Chen Tan. Towards ultra-low-power neuromorphic speech enhancement with spiking-fullsubnet. *arXiv preprint arXiv:2410.04785*, 2024.

[He *et al.*, 2024] Linxuan He, Yunhui Xu, Weihua He, Yihan Lin, Yang Tian, Yujie Wu, Wenhui Wang, Ziyang Zhang, Junwei Han, Yonghong Tian, et al. Network model with internal complexity bridges artificial intelligence and neuroscience. *Nature Computational Science*, 4(8):584–599, 2024.

[Krizhevsky and Hinton, 2009] Alex Krizhevsky and Geoffrey Hinton. Learning multiple layers of features from tiny images. Technical Report 0, University of Toronto, Toronto, Ontario, 2009.

[LeCun *et al.*, 1998] Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998.

[Lee *et al.*, 2019] Min-Ho Lee, O-Yeon Kwon, Yong-Jeong Kim, Hong-Kyung Kim, Young-Eun Lee, John Williamson, Siamac Fazli, and Seong-Whan Lee. Eeg dataset and openbmi toolbox for three bci paradigms: an investigation into bci illiteracy. *GigaScience*, 8(5):giz002, 01 2019.

[Li *et al.*, 2017] Hongmin Li, Hanchao Liu, Xiangyang Ji, Guoqi Li, and Luping Shi. Cifar10-dvs: an event-stream dataset for object classification. *Frontiers in Neuroscience*, 11:244131, 2017.

[Ma *et al.*, 2024] De Ma, Xiaofei Jin, Shichun Sun, Yitao Li, Xundong Wu, Youneng Hu, Fangchao Yang, Huajin Tang, Xiaolei Zhu, Peng Lin, et al. Darwin3: a large-scale neuromorphic chip with a novel isa and on-chip learning. *National Science Review*, 11(5):nwae102, 2024.

[Maass, 1997] Wolfgang Maass. Networks of spiking neurons: the third generation of neural network models. *Neural Networks*, 10(9):1659–1671, 1997.

[Pei *et al.*, 2019] Jing Pei, Lei Deng, Sen Song, Mingguo Zhao, Youhui Zhang, Shuang Wu, Guanrui Wang, Zhe Zou, Zhenzhi Wu, Wei He, et al. Towards artificial general intelligence with hybrid tianjic chip architecture. *Nature*, 572(7767):106–111, 2019.

[Qian *et al.*, 2021] Xinyuan Qian, Bidisha Sharma, Amine El Abridi, and Haizhou Li. Sloclas: A database for

joint sound localization and classification. *arXiv preprint arXiv:2108.02539*, 2021.

[Stan and Rhodes, 2024] Matei-Ioan Stan and Oliver Rhodes. Learning long sequences in spiking neural networks. *Scientific Reports*, 14(1):21957, 2024.

[Sun *et al.*, 2024] Pengfei Sun, Jibin Wu, Malu Zhang, Paul Devos, and Dick Botteldooren. Delayed memory unit: Modeling temporal dependency through delay gate. *IEEE Transactions on Neural Networks and Learning Systems*, 2024.

[Tan *et al.*, 2022] Ganchao Tan, Yang Wang, Han Han, Yang Cao, Feng Wu, and Zheng-Jun Zha. Multi-grained spatio-temporal features perceived network for event-based lip-reading. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 20094–20103, 2022.

[Tay *et al.*, 2020] Yi Tay, Mostafa Dehghani, Samira Abnar, Yikang Shen, Dara Bahri, Philip Pham, Jinfeng Rao, Liu Yang, Sebastian Ruder, and Donald Metzler. Long range arena: A benchmark for efficient transformers. *arXiv preprint arXiv:2011.04006*, 2020.

[Timcheck *et al.*, 2023] Jonathan Timcheck, Sumit Bam Shrestha, Daniel Ben Dayan Rubin, Adam Kupryjanow, Garrick Orchard, Lukasz Pindor, Timothy Shea, and Mike Davies. The intel neuromorphic dns challenge. *Neuromorphic Computing and Engineering*, 3(3):034005, 2023.

[Wang *et al.*, 2022] Yanxiang Wang, Xian Zhang, Yiran Shen, Bowen Du, Guangrong Zhao, Lizhen Cui, and Hongkai Wen. Event-stream representation for human gaits identification using deep neural networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(7):3436–3449, 2022.

[Wang *et al.*, 2024] Xiao Wang, Zongzhen Wu, Bo Jiang, Zhimin Bao, Lin Zhu, Guoqi Li, Yaowei Wang, and Yonghong Tian. Hardvs: Revisiting human activity recognition with dynamic vision sensors. *Proceedings of the AAAI Conference on Artificial Intelligence*, 38(6):5615–5623, Mar. 2024.

[Warden, 2018] Pete Warden. Speech commands: A dataset for limited-vocabulary speech recognition. *arXiv preprint arXiv:1804.03209*, 2018.

[Weiss, 2019] Gary M Weiss. Wisdm smartphone and smartwatch activity and biometrics dataset. *UCI Machine Learning Repository: WISDM Smartphone and Smartwatch Activity and Biometrics Dataset Data Set*, 7:133190–133202, 2019.

[Wu *et al.*, 2018] Yujie Wu, Lei Deng, Guoqi Li, Jun Zhu, and Luping Shi. Spatio-temporal backpropagation for training high-performance spiking neural networks. *Frontiers in Neuroscience*, 12:331, 2018.

[Yang *et al.*, 2024] Zheyu Yang, Taoyi Wang, Yihan Lin, Yuguo Chen, Hui Zeng, Jing Pei, Jiazheng Wang, Xue Liu, Yichun Zhou, Jianqiang Zhang, et al. A vision chip with complementary pathways for open-world sensing. *Nature*, 629(8014):1027–1033, 2024.

[Yao *et al.*, 2024] Man Yao, Jiakui Hu, Zhaokun Zhou, Li Yuan, Yonghong Tian, Bo Xu, and Guoqi Li. Spike-driven transformer. *Advances in neural information processing systems*, 36, 2024.

[Yin *et al.*, 2021] Bojian Yin, Federico Corradi, and Sander M Bohté. Accurate and efficient time-domain classification with adaptive spiking recurrent neural networks. *Nature Machine Intelligence*, 3(10):905–913, 2021.

[Yin *et al.*, 2023] Bojian Yin, Federico Corradi, and Sander M Bohté. Accurate online training of dynamical spiking neural networks through forward propagation through time. *Nature Machine Intelligence*, 5(5):518–527, 2023.

[Zhang *et al.*, 2024] Shimin Zhang, Qu Yang, Chenxiang Ma, Jibin Wu, Haizhou Li, and Kay Chen Tan. Tc-lif: A two-compartment spiking neuron model for long-term sequential modelling. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pages 16838–16847, 2024.

[Zheng *et al.*, 2024] Hanle Zheng, Zhong Zheng, Rui Hu, Bo Xiao, Yujie Wu, Fangwen Yu, Xue Liu, Guoqi Li, and Lei Deng. Temporal dendritic heterogeneity incorporated with spiking neural networks for learning multi-timescale dynamics. *Nature Communications*, 15(1):277, 2024.

[Zhou *et al.*, 2024] Shibo Zhou, Bo Yang, Mengwen Yuan, Runhao Jiang, Rui Yan, Gang Pan, and Huajin Tang. Enhancing snn-based spatio-temporal learning: A benchmark dataset and cross-modality attention model. *Neural Networks*, 180:106677, 2024.