A SPIKING NEURAL NETWORK LEARNING RULE INTEGRATING MULTI-SCALE ELIGIBILITY TRACES AND STEADY-STATE ADAPTATION

Anonymous authors
Paper under double-blind review

ABSTRACT

Spiking neural networks (SNNs) hold substantial promise for temporal modelling and neuromorphic computing owing to their event-driven computation and binary spike communication. Existing training methods, however, still face difficult temporal credit assignment, rigid surrogate gradients, and limited adaptive capacity. Inspired by spike-timing-dependent plasticity and meta-plasticity, we propose multi-scale eligibility traces with meta-plasticity and steady-state adaptation (MEMSA), a spatiotemporally local learning rule. MEMSA combines multi-scale eligibility traces for multi-resolution temporal dependencies, an activity-dependent meta-plastic surrogate gradient, and steady-state threshold adaptation that stabilizes firing. A binary-matrix local learning signal removes global error backpropagation. MEMSA preserves linear memory complexity $O(Ln)$ and computational complexity $O(LCn)$ up to constant factors. On CIFAR10-DVS, MEMSA improves upon the temporally and spatially local learning rule baseline and achieves performance comparable to standard Backpropagation Through Time within overlapping standard deviations (76.64% versus 76.40%). Competitive results are also obtained on CIFAR-100 and IBM DVS Gesture. These results indicate that fully local learning can achieve performance comparable to global backpropagation on the evaluated event-based benchmarks.

1 INTRODUCTION

The ubiquity of edge devices has intensified the demand for energy-efficient machine learning algorithms under resource-constrained conditions (Roy et al., 2019; Eshraghian et al., 2023). Spiking neural networks (SNNs), characterized by event-driven computation and binary spike communication, are inherently suited for high-efficiency neuromorphic hardware implementations. However, training SNNs on such platforms is hindered by the spatiotemporal credit-assignment problem. The standard backpropagation through time (BPTT) algorithm (Werbos, 1990) incurs $O(TLn^2)$ time complexity and $O(TLn)$ memory complexity, with updates depending on future states. Such global dependencies impose prohibitive memory and communication overheads.

To mitigate these limitations, various methods have been explored to achieve local credit assignment. For spatial credit assignment, methods such as Feedback Alignment (Lillicrap et al., 2016) and Direct Feedback Alignment (DFA) (Crafton et al., 2019) utilize random matrices to avoid exact backpropagation. Other biologically plausible alternatives replace backward passes with forward executions, such as the Forward-Forward algorithm (Hinton, 2022). More recently, the Local Learning rule inspired by neural activity Synchronization has been introduced using fixed periodic basis vectors for layer-wise training, achieving performance competitive with backpropagation on larger datasets (Apolinario et al., 2025a).

For temporal credit assignment in SNNs, three-factor learning rules (Frémaux & Gerstner, 2016) with eligibility traces (Shouval & Kirkwood, 2025) provide biologically plausible solutions. Eligibility propagation addresses temporal credit assignment in recurrent SNNs through eligibility traces (Bellec et al., 2020). Other frameworks combine eligibility traces with direct random target projection to achieve spatiotemporal locality (Ortner et al., 2023; Frenkel et al., 2021), although their evaluation has mainly focused on shallow models. In contrast, the spike-timing-dependent plastic-

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

ity (STDP)-inspired Temporal Local Learning Rule (S-TLLR) achieves linear memory complexity $O(Ln)$, but its spatial credit assignment still requires cross-layer error propagation (Apolinario & Roy, 2024). Building upon S-TLLR, the Temporally and Spatially Local Learning Rule (TESS) achieves full spatiotemporal locality through fused traces and local learning signals (Apolinario et al., 2025b). However, its single-scale trace, static surrogate gradient, and fixed threshold limit its adaptive capacity.

To overcome these constraints, Multi-scale Eligibility Traces with Meta-plasticity and Steady-state Adaptation (MEMSA) is proposed as a scalable spatiotemporally local SNN learning rule for resource-constrained scenarios. K parallel traces are introduced to construct Multi-scale Eligibility Traces (MET), allowing temporal dependencies at different resolutions to be represented (Gerstner et al., 2018). Surrogate-gradient support is adapted according to neuronal activity through the Meta-plasticity Spike Gradient (MSG), following established principles of meta-plasticity (Abraham & Bear, 1996; Citri & Malenka, 2008). Firing activity is stabilized through closed-loop Steady-state Threshold Adaptation (STA) (Liu & Minch, 2000). The local signal generation mechanism of TESS is retained to eliminate temporal unrolling and cross-layer error propagation.

The contributions of this work are summarized as follows:

1. A fully local SNN learning rule is developed using multi-scale eligibility traces, activity-dependent meta-plasticity, and threshold homeostasis. All adaptive states remain neuron- or synapse-local, supporting energy-efficient edge learning.
2. MEMSA maintains $O(Ln)$ memory complexity and $O(LCn)$ per-time-step computational complexity. The required state storage remains independent of the sequence length, while the learning capacity is extended through MET, MSG, and STA.
3. Evaluations on CIFAR-100, IBM DVS Gesture, and CIFAR10-DVS demonstrate competitive performance across static and event-based benchmarks. In particular, MEMSA improves upon TESS by 1.64 percentage points on CIFAR10-DVS and achieves performance comparable to BPTT while retaining the same asymptotic complexity as TESS.

2 METHODOLOGY

This section presents the methodological foundation of MEMSA. The leaky integrate-and-fire neuron model and gradient-based optimization are first introduced, followed by the three-factor learning framework. The three components of MEMSA, namely multi-scale eligibility traces (MET), the meta-plastic spike gradient (MSG), and steady-state threshold adaptation (STA), are then described. Finally, the complete local update rule and its computational and memory complexities are presented. As illustrated in Figure 1, MET captures temporal dependencies at multiple resolutions through parallel trace paths, MSG adjusts the surrogate-gradient window according to neuronal activity, and STA regulates firing activity through adaptive threshold control.

2.1 NEURON MODEL AND GRADIENT OPTIMIZATION

A discrete-time leaky integrate-and-fire neuron is defined as

$$u_i^{(l)}[t] = \gamma \left(u_i^{(l)}[t-1] - v_{\mathrm{th}} o_i^{(l)}[t-1] \right) + \sum_j W_{ij}^{(l)} o_j^{(l-1)}[t], \tag{1}$$

where $\gamma = \exp(-dt/\tau)$ is the leak factor, $W_{ij}^{(l)}$ is the synaptic weight, and v_{th} is the firing threshold. The binary output spike is

$$o_i^{(l)}[t] = \Theta\left(u_i^{(l)}[t] - v_{\mathrm{th}} \right), \tag{2}$$

where $o_i^{(l)}[t] \in \{0, 1\}$ and $\Theta(\cdot)$ denotes the Heaviside step function.

For an L-layer network unfolded over T time steps, the BPTT gradient is

$$\frac{\partial \mathcal{L}}{\partial W^{(l)}} = \sum_{t=1}^{T} \frac{\partial \mathcal{L}}{\partial u^{(l)}[t]} \frac{\partial u^{(l)}[t]}{\partial W^{(l)}}. \tag{3}$$

Because $\partial \mathcal{L} / \partial u^{(l)}[t]$ depends on future recurrent states, BPTT requires temporal unrolling and does not satisfy temporal locality.

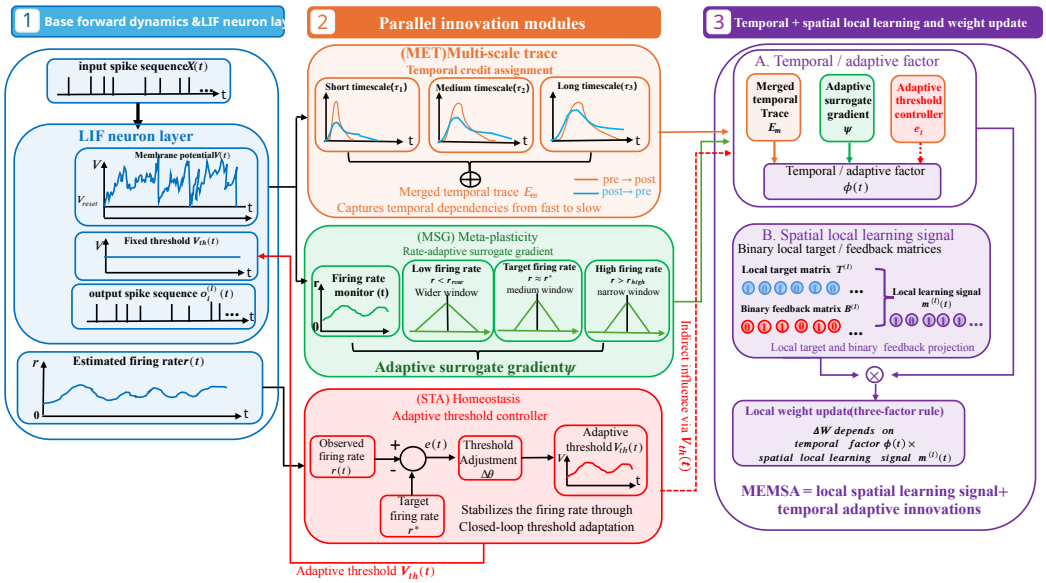

Figure 1: Overview of the MEMSA framework. Multi-scale eligibility traces (MET), the meta-plastic spike gradient (MSG), and steady-state threshold adaptation (STA) provide temporally adaptive factors, while a fixed binary projection produces spatially local learning signals.

2.2 Three-factor learning rules

A three-factor rule combines pre-synaptic activity, post-synaptic activity, and a modulatory learning signal. A generic eligibility trace is defined as

$$e_{ij}^{(l)}[t] = \beta e_{ij}^{(l)}[t-1] + f\left(o_i^{(l)}[t]\right) g\left(o_j^{(l-1)}[t]\right), \tag{4}$$

where $\beta < 1$ is the trace-decay factor and $f(\cdot)g(\cdot)$ captures the immediate pre–post correlation. The corresponding weight update is

$$\Delta W_{ij} = \sum_t m_i[t] e_{ij}^{(l)}[t], \tag{5}$$

where $m_i[t]$ is the modulatory signal. The eligibility trace stores recent correlations, while the modulatory signal determines how strongly they contribute to learning. This factorization allows synaptic evidence to be retained locally without storing the complete temporal computation graph.

2.3 Multi-scale eligibility traces

Inspired by eligibility traces that operate across behavioral time scales (Gerstner et al., 2018), MET maintains K causal and non-causal trace paths with different decay factors:

$$q_k^{(l)}[t] = \lambda_k q_k^{(l)}[t-1] + o^{(l-1)}[t], \qquad \lambda_k = \exp(-1/\tau_k), \tag{6}$$

$$h_k^{(l)}[t] = \mu_k h_k^{(l)}[t-1] + \Psi\left(u^{(l)}[t-1]\right), \qquad \mu_k = \exp(-1/\hat{\tau}_k), \tag{7}$$

where $\Psi(\cdot)$ denotes the surrogate gradient. In the evaluated configuration, $K = 3$ and $\tau_k \in \{2, 8, 32\}$ are used to represent short-, medium-, and long-range temporal dependencies.

The parallel traces are fused using normalized Gaussian weights:

$$w_k = \frac{\exp\left[-(k-(K+1)/2)^2/(2\sigma^2)\right]}{\sum_{j=1}^{K} \exp\left[-(j-(K+1)/2)^2/(2\sigma^2)\right]}, \qquad \sigma = 0.75. \tag{8}$$

The initial fused traces are $q_{\text{mix}}^{(l)}[t] = \sum_k w_k q_k^{(l)}[t]$ and $h_{\text{mix}}^{(l)}[t] = \sum_k w_k h_k^{(l)}[t]$. Because traces with different decay factors accumulate energy at different rates, an energy correction

$g = \mathrm{clamp}(E_{\mathrm{ref}}/\|q_{\mathrm{mix}}\|, 1 \pm \epsilon)$ is applied. A conflict-aware blending coefficient further reduces the multi-scale contribution when its direction differs substantially from the fast reference trace. The resulting energy-normalized traces are denoted by $q^{(l)}_{\mathrm{final}}[t]$ and $h^{(l)}_{\mathrm{final}}[t]$.

The causal and non-causal eligibility components are then defined as

$$e^{(l)}_{\mathrm{pre}}[t] = \alpha_{\mathrm{pre}} \Psi\Big(u^{(l)}[t]\Big) \otimes q^{(l)}_{\mathrm{final}}[t], \tag{9}$$

$$e^{(l)}_{\mathrm{post}}[t] = \alpha_{\mathrm{post}} h^{(l)}_{\mathrm{final}}[t] \otimes o^{(l-1)}[t]. \tag{10}$$

For each scale, $\lambda_k > \mu_k$ makes the causal trace persist longer than the non-causal trace, producing an asymmetric STDP-like temporal window. The shortest trace responds rapidly to transient events, the intermediate trace represents patterns spanning several simulation steps, and the longest trace preserves slowly varying temporal context. Energy matching and conflict-aware blending prevent the slow traces from dominating the fused update.

2.4 META-PLASTIC SPIKE GRADIENT

Meta-plasticity describes changes in the capacity of a synapse or neuron for subsequent plasticity (Abraham & Bear, 1996; Citri & Malenka, 2008). Based on this principle, MSG adapts the surrogate-gradient window using a smoothed firing-rate estimate:

$$\bar{r}^{(l)}_i[t] = \beta_r \bar{r}^{(l)}_i[t-1] + (1-\beta_r) o^{(l)}_i[t]. \tag{11}$$

The activity-dependent width is

$$\rho^{(l)}_i = \rho_{\mathrm{base}} \exp\Big(\kappa(r_{\mathrm{target}} - \bar{r}^{(l)}_i)\Big), \tag{12}$$

and the adaptive surrogate derivative is

$$\Psi_{\mathrm{adapt}}(u, \rho) = 0.3 \max\left(1 - \frac{|u - v_{\mathrm{th}}|}{\rho w_{\mathrm{base}}}, 0\right), \tag{13}$$

where w_{base} is the baseline window width. When $\bar{r}^{(l)}_i < r_{\mathrm{target}}$, the window is widened so that membrane potentials farther from the threshold can receive learning signals. When $\bar{r}^{(l)}_i > r_{\mathrm{target}}$, the window is narrowed to restrict updates in highly active neurons. The local error and eligibility trace continue to determine the update direction; MSG changes only the support of the surrogate derivative.

2.5 STEADY-STATE THRESHOLD ADAPTATION

Following the homeostatic principle that neuronal excitability can be adjusted to stabilize activity (Liu & Minch, 2000), STA decomposes the firing threshold into a base value and an adaptive offset:

$$v^{(l)}_{\mathrm{th},i}[t] = v_{\mathrm{th,base}} + \delta^{(l)}_i[t]. \tag{14}$$

The short-term offset follows

$$\delta^{(l)}_i[t] = \beta_\delta \delta^{(l)}_i[t-1] + \eta_\delta \left(o^{(l)}_i[t-1] - r_{\mathrm{target}}\right), \tag{15}$$

where β_δ controls the decay and η_δ is the adaptation rate. Long-term homeostatic regulation uses a dead zone and an asymmetric response:

$$\Delta_{\mathrm{homeo}} = \eta_h \left[\max(\mathrm{err} - \theta_{\mathrm{dead}}, 0) - \alpha_{\mathrm{down}} \max(-\mathrm{err} - \theta_{\mathrm{dead}}, 0)\right]. \tag{16}$$

The update is clipped to $|\Delta_{\mathrm{homeo}}| \le 10^{-4}$, α_{down} is set to 0.25, and the threshold is bounded within $[0.3, 1.2]$. The dead zone suppresses small oscillations, clipping limits individual threshold changes, and the asymmetric response penalizes excessive activity more strongly than insufficient activity. The target firing rate therefore acts as a controller reference rather than a hard equality constraint.

Coupled adaptation. MET preserves temporal evidence at multiple resolutions, MSG regulates whether that evidence contributes to learning, and STA stabilizes the activity resulting from the update. The three mechanisms therefore operate at complementary stages of the same local learning process without introducing future-state dependencies or cross-layer error propagation.

Table 1: Structural and complexity comparison of SNN learning rules.

Method	Memory	Computation/step	Temporal	Spatial
BPTT	$O(TLn)$	$O(Ln^2)$	No	No
DFA	$O(CLn)$	$O(Ln^2)$	No	Yes
S-TLLR	$O(Ln)$	$O(Ln^2)$	Yes	No
TESS	$O(Ln)$	$O(LCn)$	Yes	Yes
MEMSA	$O(Ln)$	$O(LCn)$	Yes	Yes

2.6 Final update and complexity

The complete local update is

$$\Delta W^{(l)}[t] = \alpha_{\text{pre}} \left(m^{(l)}[t] \odot \Psi_{\text{adapt}}(u^{(l)}[t]) \right) \otimes q^{(l)}_{\text{final}}[t] + \alpha_{\text{post}} \left(m^{(l)}[t] \odot h^{(l)}_{\text{final}}[t] \right) \otimes o^{(l-1)}[t], \tag{17}$$

where

$$m^{(l)}[t] = B^{(l)\top} \left(\text{softmax}(B^{(l)} o^{(l)}[t]) - y^* \right) \tag{18}$$

is generated through a fixed binary projection $B^{(l)} \in \mathbb{R}^{C \times n^{(l)}}$, and y^* is the target label vector. The learning signal is generated independently at each layer, so no cross-layer error propagation is required.

Table 1 summarizes the structural properties and asymptotic complexities of the evaluated SNN learning rules.

BPTT stores $O(TLn)$ hidden states, whereas MEMSA maintains only the current neuron states and eligibility traces. With fixed K, the trace storage $2K \sum_l n^{(l)}$ remains $O(Ln)$ and is independent of the sequence length T. The local projection cost over a complete sequence is

$$\text{MAC}_{\text{MEMSA}} = \sum_{l=1}^{L} (T - t_l)\, 2n^{(l)} C, \tag{19}$$

where t_l denotes the first update time of layer l. The resulting sequence-level computation is $O(TLCn)$, while the per-time-step computation is $O(LCn)$. Because T is fixed by the input encoding protocol for each benchmark, Table 1 reports per-time-step complexity. Binary projection avoids spatial backpropagation and preserves linear scaling with network width.

3 Experimental Results

This section evaluates MEMSA on static and event-based benchmarks. The datasets, network architecture, training protocol, and fixed hyperparameters are first described. Ablation studies are then used to examine the contributions of MET, MSG, and STA. Finally, MEMSA is compared with representative global and local SNN learning methods.

3.1 Experimental settings

Experiments are conducted on CIFAR-100 (Krizhevsky, 2009), IBM DVS Gesture (Amir et al., 2017), and CIFAR10-DVS (Li et al., 2017). CIFAR-100 uses $T = 6$ repeated presentations, DVS Gesture uses 20 frames at 75 ms per frame, and CIFAR10-DVS uses 10 frames. Data augmentation consists of random cropping with four-pixel zero padding, horizontal flipping, and cutout. A VGG-9 network with global average pooling is trained for 200 epochs using Adam with an initial learning rate of 10^{-3}, which is halved after a five-epoch validation plateau. Results are reported as the mean $\pm$ standard deviation over three random seeds.

The fixed configuration is $K = 3$, $\tau \in \{2, 8, 32\}$, $\beta_r = 0.98$, $r_{\text{target}} = 0.1$, $\rho_{\text{base}} = 1.0$, $\eta_h = 7 \times 10^{-4}$, and $\beta_\delta = 0.95$. STA is linearly annealed during late training and subsequently frozen. The same backbone, training schedule, and evaluation protocol are used for matched comparisons.

Table 2: Component-wise ablation across benchmark datasets (%).

Configuration	CIFAR10-DVS	CIFAR-100	DVS Gesture
Baseline (TESS)	75.00 ± 0.65	70.00 ± 0.34	98.56 ± 0.31
+MET	75.93 ± 0.85	69.94 ± 0.36	98.56 ± 0.31
+MSG	75.67 ± 1.56	70.13 ± 0.39	98.61 ± 0.22
+STA	76.03 ± 0.38	70.00 ± 0.23	98.61 ± 0.22
MEMSA (full)	$\mathbf{76.64 \pm 0.64}$	70.00 ± 0.45	$\mathbf{98.74 \pm 0.58}$

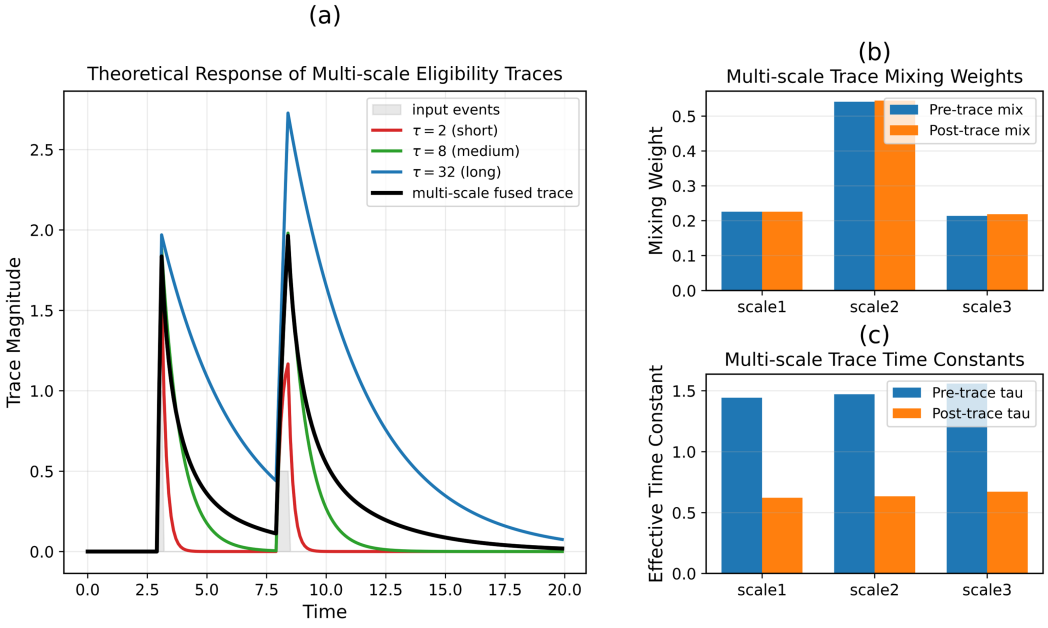

Figure 2: Validation of the multi-scale eligibility-trace mechanism: (a) responses at three time scales and their fusion, (b) Gaussian mixing weights, and (c) effective causal and non-causal time constants.

3.2 Ablation Studies

Table 2 reports the component-wise ablation results for MET, MSG, and STA. Each module is first evaluated individually on top of the TESS baseline, and the complete configuration contains all three modules. The same backbone, optimizer, training schedule, and number of random seeds are retained across the configurations so that the reported differences are associated with the introduced adaptive mechanisms.

Contribution of MET. MET increases CIFAR10-DVS accuracy from 75.00% to 75.93%, while producing little change on CIFAR-100 and DVS Gesture. Figure 2 shows that the short trace responds rapidly and decays quickly, the intermediate trace follows patterns spanning several simulation steps, and the long trace preserves more persistent temporal context. The Gaussian weights emphasize the intermediate scale while retaining the short and long components. The effective causal time constants are larger than the corresponding non-causal values, which is consistent with the asymmetric temporal window used by the update rule. On CIFAR-100, the same static image is repeatedly presented over six time steps, so the parallel traces receive highly redundant inputs. The larger change on CIFAR10-DVS is therefore consistent with the greater temporal variation of event-stream inputs.

Contribution of MSG. MSG achieves $75.67 \pm 1.56\%$ on CIFAR10-DVS and $70.13 \pm 0.39\%$ on CIFAR-100. Figure 3 illustrates the three activity-dependent states described in the original analysis. Below the target firing rate, the surrogate-gradient window is broadened so that membrane

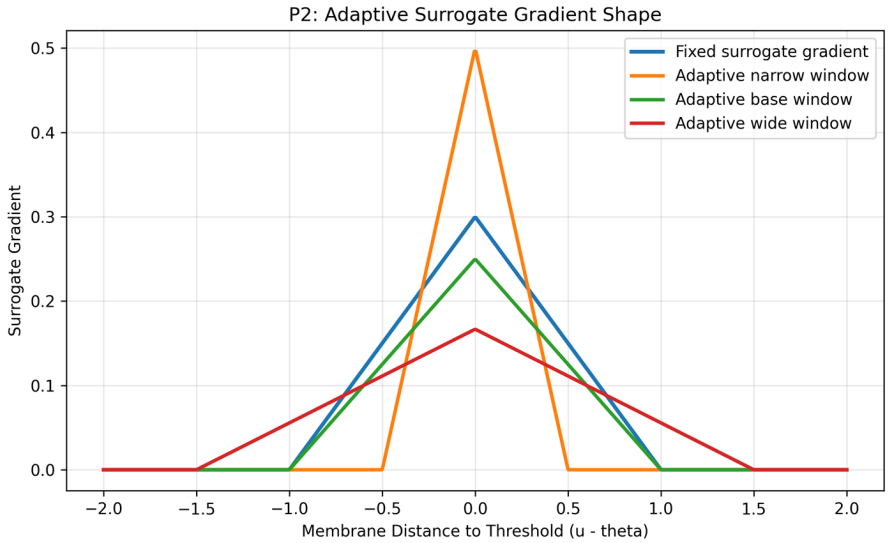

Figure 3: Adaptive surrogate-gradient shapes under different firing rates.

potentials farther from the threshold can contribute to learning. Near the target rate, the baseline profile is retained. Above the target rate, the window is narrowed to restrict updates in highly active neurons. The update direction remains determined by the local error and eligibility trace. The larger standard deviation of the MSG-only configuration on CIFAR10-DVS suggests greater sensitivity to random initialization, whereas lower variation is observed in the complete configuration.

Contribution of STA. STA provides the largest individual improvement on CIFAR10-DVS, reaching $76.03 \pm 0.38\%$, compared with $75.00 \pm 0.65\%$ for the baseline. Smaller changes are observed on CIFAR-100 and DVS Gesture. As shown in Figure 4, the average firing rate of MEMSA decreases during the early adaptation period and subsequently remains below the baseline trajectory. This behavior follows the homeostatic principle used in the original design: neuronal excitability is adjusted when activity departs from its reference level. The firing rate does not converge exactly to r_{target} because the controller includes a dead zone, asymmetric updates, clipping, and bounded offsets. The target rate therefore serves as a reference rather than a strict equality constraint.

Combined configuration and gradient flow. When all three mechanisms are combined, MEMSA reaches $76.64 \pm 0.64\%$ on CIFAR10-DVS and $98.74 \pm 0.58\%$ on DVS Gesture. On CIFAR-100, the complete model obtains $70.00 \pm 0.45\%$, matching the baseline mean. The improvement is therefore most apparent on event-based inputs. MET retains temporal evidence at several resolutions, MSG adjusts the surrogate-gradient support according to activity, and STA regulates the firing activity produced during learning. Figure 5 compares the gradient norms of the baseline and MEMSA across network layers. In both methods, the values are small in the shallow layers and become larger in the deeper layers. The gradient norms produced by MEMSA remain of the same order as those of the baseline, showing that measurable learning signals are retained in the deeper layers. Together, the ablation results show that the three mechanisms provide complementary forms of temporal representation and activity adaptation within the same local-learning framework.

3.3 Comparison with State-of-the-Art Methods

Table 3 compares MEMSA with representative global and local SNN learning methods under the same backbone and evaluation protocol.

CIFAR10-DVS. MEMSA achieves $76.64 \pm 0.64\%$, which is 1.64 percentage points higher than the fully local TESS baseline at $75.00 \pm 0.65\%$. BPTT obtains $76.40 \pm 0.66\%$. Because the standard deviations of MEMSA and BPTT overlap, the two methods are regarded as providing comparable

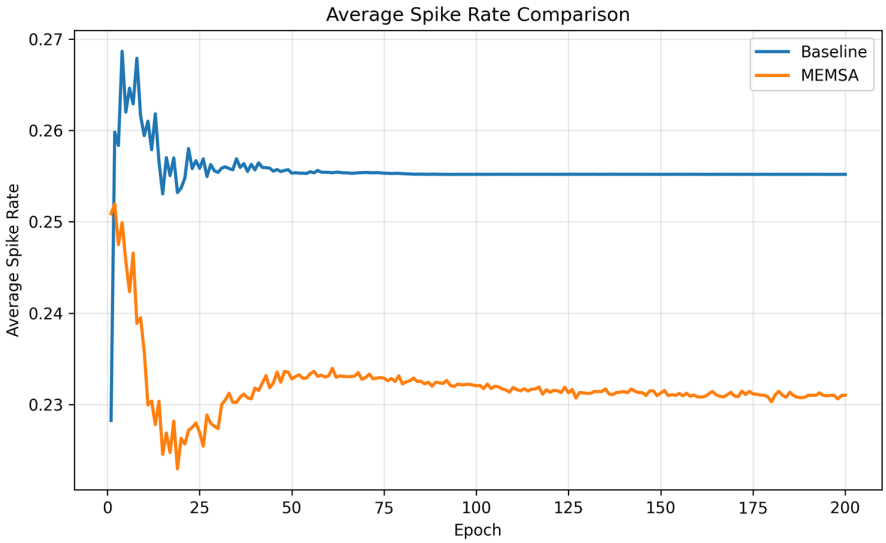

Figure 4: Firing-rate balancing effect of steady-state threshold adaptation.

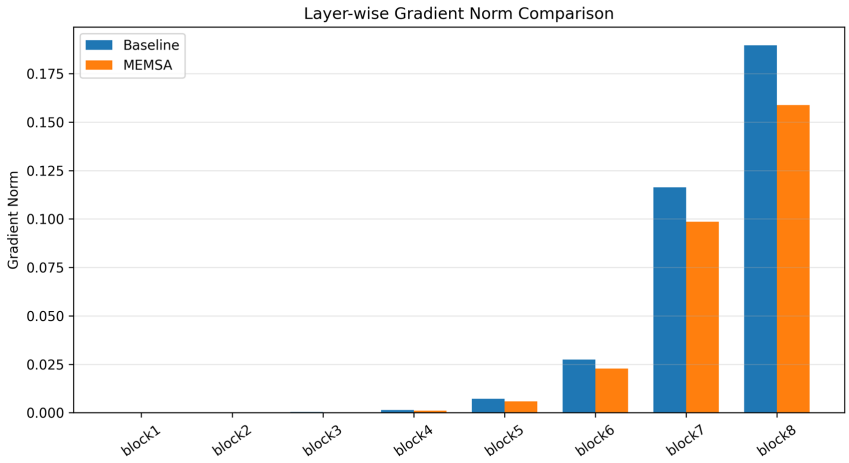

Figure 5: Per-layer gradient-norm comparison between MEMSA and the baseline.

performance on this benchmark. Unlike BPTT, MEMSA does not require temporal unrolling or cross-layer error propagation.

IBM DVS Gesture. MEMSA reaches $98.74 \pm 0.58\%$, compared with $98.56 \pm 0.31\%$ for TESS and $97.95 \pm 0.68\%$ for BPTT. These results indicate that the proposed mechanisms maintain competitive performance on a temporally structured gesture-recognition task. Since the reported variation overlaps and all three accuracies are close to saturation, the small differences in their mean values should not be interpreted as conclusive superiority.

CIFAR-100. MEMSA obtains $70.00 \pm 0.45\%$, matching the mean accuracy of TESS and exceeding the reported BPTT mean of $69.28 \pm 0.37\%$. OTTT achieves the highest reported result of $71.05 \pm 0.04\%$, but it provides temporal locality rather than full spatiotemporal locality. The limited change relative to TESS is consistent with the repeated static input used for CIFAR-100, which contains less temporal variation than the event-based datasets.

Table 3: Comparison with representative SNN learning rules (%).

Method	Locality	CIFAR10-DVS	DVS Gesture	CIFAR-100
BPTT	Non-local	76.40 ± 0.66	97.95 ± 0.68	69.28 ± 0.37
S-TLLR	Temporal	75.14 ± 1.37	98.48 ± 0.37	68.00 ± 0.71
OTTT	Temporal	76.27 ± 0.05	96.88	71.05 ± 0.04
DECOLLE	Full	–	95.54 ± 0.16	–
TESS	Full	75.00 ± 0.65	98.56 ± 0.31	70.00 ± 0.34
MEMSA	Full	$\mathbf{76.64 \pm 0.64}$	$\mathbf{98.74 \pm 0.58}$	70.00 ± 0.45

Overall, the improvement over TESS is most apparent on CIFAR10-DVS, whereas the differences on IBM DVS Gesture and CIFAR-100 are smaller. Across the three benchmarks, MEMSA maintains fully local credit assignment and achieves performance comparable to the evaluated global and temporally local methods.

4 Conclusion

This paper proposes MEMSA, an extension of the STDP rule with three mechanisms (MET, MSG, and STA). Multi-resolution temporal information is represented through parallel traces, surrogate-gradient support is adjusted according to neuronal activity, and firing activity is regulated through threshold adaptation. Across the evaluated benchmarks, MEMSA improves upon TESS on CIFAR10-DVS and achieves performance comparable to BPTT within overlapping standard deviations, while maintaining local updates and linear memory scaling with network size. Future work will investigate automatic selection of temporal scales and time constants, asynchronous event-driven computation, and validation on neuromorphic hardware.

AI use statement

Generative AI tools are not used for this work.

Ethics statement

This work uses established public vision benchmarks and does not involve human-subject data collection. Potential applications include low-power sensing and edge intelligence. Practical deployment should account for privacy, security, dataset bias, and the reliability of decisions produced by resource-constrained learning systems.

References

Wickliffe C. Abraham and Mark F. Bear. Metaplasticity: The plasticity of synaptic plasticity. *Trends in Neurosciences*, 19(4):126–130, 1996.

Arnon Amir et al. A low power, fully event-based gesture recognition system. In *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 7388–7397, 2017.

Mariano P. E. Apolinario and Kaushik Roy. S-tllr: Stdp-inspired temporal local learning rule for spiking neural networks. *arXiv preprint arXiv:2306.15220v4*, 2024.

Mariano P. E. Apolinario, Arindam Roy, and Kaushik Roy. Lls: Local learning rule for deep neural networks inspired by neural activity synchronization. In *Winter Conference on Applications of Computer Vision*, pp. 7807–7816, 2025a.

Mariano P. E. Apolinario, Kaushik Roy, and Charlotte Frenkel. Tess: A scalable temporally and spatially local learning rule for spiking neural networks. In *International Joint Conference on Neural Networks*, pp. 1–9, 2025b.

Guillaume Bellec et al. A solution to the learning dilemma for recurrent networks of spiking neurons. *Nature Communications*, 11:1–15, 2020.

Ami Citri and Robert C. Malenka. Synaptic plasticity: Multiple forms, functions, and mechanisms. *Neuropsychopharmacology*, 33(1):18–41, 2008.

Brian Crafton, Abhinav Parihar, Evan Gebhardt, and Arijit Raychowdhury. Direct feedback alignment with sparse connections for local learning. *Frontiers in Neuroscience*, 13(525):1–12, 2019.

Jason K. Eshraghian et al. Training spiking neural networks using lessons from deep learning. *Proceedings of the IEEE*, 111(9):1016–1054, 2023.

Nicolas Frémaux and Wulfram Gerstner. Neuromodulated spike-timing-dependent plasticity, and theory of three-factor learning rules. *Frontiers in Neural Circuits*, 9(85):1–19, 2016.

Charlotte Frenkel, Martin Lefebvre, and David Bol. Learning without feedback: Fixed random learning signals allow for feedforward training of deep neural networks. *Frontiers in Neuroscience*, 15:1–13, 2021.

Wulfram Gerstner, Marco Lehmann, Vasiliki Liakoni, Dane Corneil, and Johanni Brea. Eligibility traces and plasticity on behavioral time scales: Experimental support of neo-hebbian three-factor learning rules. *Frontiers in Neural Circuits*, 12(53):1–16, 2018.

Geoffrey Hinton. The forward-forward algorithm: Some preliminary investigations. *arXiv preprint arXiv:2212.13345*, 2022.

Alex Krizhevsky. Learning multiple layers of features from tiny images. Technical report, University of Toronto, 2009.

Hongmin Li, Hanchao Liu, Xiangyang Ji, Guoqi Li, and Luping Shi. Cifar10-dvs: An event-stream dataset for object classification. *Frontiers in Neuroscience*, 11:1–10, 2017.

Timothy P. Lillicrap, Daniel Cownden, Douglas B. Tweed, and Colin J. Akerman. Random synaptic feedback weights support error backpropagation for deep learning. *Nature Communications*, 7: 1–10, 2016.

Shih-Chii Liu and Bradley A. Minch. Homeostasis in a silicon integrate-and-fire neuron. In *Advances in Neural Information Processing Systems*, pp. 1–7, 2000.

Thomas Ortner, Luca Pes, Julian Gentinetta, Charlotte Frenkel, and Angeliki Pantazi. Online spatio-temporal learning with target projection. In *IEEE International Conference on Artificial Intelligence Circuits and Systems*, pp. 1–5, 2023.

Kaushik Roy, Akhilesh Jaiswal, and Priyadarshini Panda. Towards spike-based machine intelligence with neuromorphic computing. *Nature*, 575(7784):607–617, 2019.

Harel Z. Shouval and Alfredo Kirkwood. Eligibility traces as a synaptic substrate for learning. *Current Opinion in Neurobiology*, 91:1–8, 2025.

Paul J. Werbos. Backpropagation through time: what it does and how to do it. *Proceedings of the IEEE*, 78(10):1550–1560, 1990.