CONDENFORMER: A CONDUCTANCE-INSPIRED SPIKING TRANSFORMER WITH DENDRITIC MEMORY AND ADAPTIVE NEURONAL DYNAMICS

Anonymous authors
Paper under double-blind review

ABSTRACT

Spiking neural networks (SNNs) process information through discrete spikes, offering a promising approach to energy-efficient, brain-inspired computation. However, conventional current-based leaky integrate-and-fire (LIF) neurons face two limitations. First, weak inputs may fail to elicit spikes within a finite simulation window, causing subthreshold evidence to be lost from the emitted spike sequence. Second, their responses depend strongly on the temporal synchrony of excitatory inputs: dispersing a fixed number of spikes over time can substantially reduce the output firing rate compared with synchronized inputs. We address these limitations with ConDenFormer, a spiking Transformer inspired by conductance-dependent integration and dendritic–somatic computation. ConDenFormer integrates adaptive conductance exchange with persistent, source-specific dendritic memory. Learnable nonnegative conductances, modulated by pre-cancellation input activity and dendritic context, control how retained evidence contributes to the somatic membrane potential and its effective response timescale. Dendritic states persist across somatic spikes, preserving input history that would otherwise be discarded. We establish a dissipative exchange law, bounded driven dynamics, and a constructive separation of input sequences with identical source and temporal marginals. Across multiple benchmarks, ConDenFormer achieves 97.21% accuracy on CIFAR-10, 83.42% on CIFAR-100, 85.4% on CIFAR10-DVS, 99.0% on DVS128 Gesture, and 79.45% on ImageNet-1K with only 16.29M parameters. These results demonstrate that source-specific memory and adaptive conductance can improve the preservation and utilization of temporal evidence in spiking visual computation.

1 INTRODUCTION

Spiking neural networks (SNNs) process information through discrete spikes and evolving neuronal states, offering a biologically inspired route to energy-efficient computation through event-driven communication. Advances in surrogate-gradient learning and membrane-level residual connections have made deep SNNs increasingly effective for visual recognition (Neftci et al., 2019; Hu et al., 2025). Spiking Transformers further extend this progress through spiking self-attention, hierarchical token mixing, and spatiotemporal attention (Zhou et al., 2023; Yao et al., 2023; Zhou et al., 2024; Wang et al., 2023; 2025b). These developments shift attention from whether SNNs can achieve competitive recognition performance to a more fundamental question of neuronal computation: *how can visual evidence be preserved when it passes through spike-based computation within a short simulation window?*

This question is particularly important for preserving fine spatial detail in visual recognition. Fang et al. (2025) report pronounced high-frequency information loss in the SNNs and show that max-pooling-based protection can mitigate this loss in both residual and Transformer architectures. Other approaches, including wavelet processing, multi-scale attention fusion, active predictive filtering, and lateral inhibition, have also been proposed to improve spiking visual representations (Fang et al., 2024; Hua et al., 2025; Xie et al., 2026; Zheng et al., 2025). These studies demonstrate that preserving visual information remains an important challenge in spiking networks, but primarily address it through spatial operators or feature-processing mechanisms. Yet before such mechanisms can

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.

exploit a feature, its evidence must first be retained by the underlying neuronal dynamics. We therefore investigate two obstacles that arise from the thresholding and temporal integration dynamics of current-based leaky integrate-and-fire (LIF) neurons.

Finite-time gating. A LIF neuron starting at rest needs sufficient integrated input to cross its firing threshold before the simulation ends. Consequently, a subthreshold input that remains meaningful in an analog feature representation may produce no output spikes. This finite-time gate can be particularly restrictive when local convolution generates low-amplitude responses to fine spatial detail. Although natural-image statistics tend to concentrate less absolute power at fine spatial scales (Ruderman, 1994), the actual drive reaching a neuron is also determined by learned filters and normalization. Thus, weak but potentially informative evidence can remain subthreshold throughout the simulation and disappear from the emitted spike sequence before reaching deeper layers.

Timing-dependent transfer. Even when a neuron emits spikes, their downstream transmission depends on when they arrive. Synchronized excitatory inputs reinforce one another before substantial membrane leakage occurs, whereas temporally dispersed inputs are integrated less effectively because earlier responses decay before later arrivals can contribute. Thus, two input sequences with the same number of spikes can produce substantially different downstream responses, potentially even differing between spiking and non-spiking regimes. This effect is relevant to visual processing because local spatial structure can induce different firing times under presentation of a static image: distinct convolutional currents may reach threshold at different latencies. Consequently, temporal structure is not merely a by-product of the spike count; it can determine how effectively information is transmitted across layers.

These two obstacles point to a common requirement: useful evidence should persist across layers, while the neuron should retain information about *when* and *from which source* that evidence arrived. A simple aggregation of input sources before integration cannot provide this distinction, because different source-specific sequences can produce the same net drive at every time step. Dendritic branches provide a natural computational organization for integrating local inputs before somatic combination (Poirazi et al., 2003; Polsky et al., 2004) and can preserve distinctions among inputs over time (Branco et al., 2010; Zheng et al., 2024). Conductance modulation offers a complementary mechanism: by regulating how strongly currents influence the membrane, conductances can alter neuronal integration even when opposing inputs cancel at the net-current level (Destexhe et al., 2003; Gütig & Sompolinsky, 2009). These observations motivate our central design principle: *retain source-specific temporal evidence and provide adaptive access to it at the soma*.

We introduce **ConDenFormer**, a spiking Transformer that implements this principle through reciprocal exchange between a LIF soma and persistent dendritic compartments. Four dendritic states retain contributions from distinct input-channel banks, while their conductance are modulated by both pre-cancellation input activity and dendritic context. Compartment capacities regulate how strongly each exchange updates the stored states. As a result, the neuron can learn when evidence from each source should influence the somatic membrane potential, yielding a source-dependent effective receptive field that varies with temporal delay. We integrate these neurons into two late spatial transitions of MaxFormer, retaining its spatial operators and membrane shortcuts while replacing the corresponding neuronal computation.

Our analysis establishes two properties of the resulting dynamics. First, we derive a dissipative exchange law that yields bounded driven state evolution. Second, we constructively show that the model can distinguish input sequences with identical source totals and identical per-step net inputs, demonstrating that source-specific memory preserves temporal distinctions that would be lost under immediate aggregation. Guided by these properties, we conduct matched neuronal comparisons, component ablations, and temporal interventions across static-image and event-based visual recognition benchmarks. Together, the analysis and experiments examine whether adaptive conductance and persistent dendritic memory improve the preservation and utilization of temporal evidence in deep SNNs.

2 Related work

Spiking vision. Direct training has substantially advanced deep SNNs through spatiotemporal backpropagation, residual learning, threshold-aware normalization, and temporal objectives (Wu et al.,

2018; Fang et al., 2021a; Zheng et al., 2021; Deng et al., 2022). Building on these developments, spiking Transformers introduce attention mechanisms and hierarchical feature extraction for visual recognition (Zhou et al., 2023; Yao et al., 2023; Zhou et al., 2024), together with spike-driven residual architectures (Chenlin Zhou, 2026). MaxFormer further combines max-pooling with local and global token mixing to mitigate the loss of fine spatial information (Fang et al., 2025). Other approaches exploit multi-scale feature fusion, predictive filtering, and lateral inhibition to improve spiking visual representations (Hua et al., 2025; Xie et al., 2026; Zheng et al., 2025). Our work retains MaxFormer's spatial hierarchy and operators while introducing new recurrent neuronal dynamics at the receiving neurons.

Neuronal and dendritic dynamics. A complementary line of work improves SNNs by enriching neuronal dynamics beyond the conventional LIF model. Learnable and heterogeneous membrane time constants adapt the temporal integration window of individual neurons (Fang et al., 2021b; Perez-Nieves et al., 2021). DH-LIF assigns inputs to heterogeneous dendritic branches and preserves their currents across somatic resets (Zheng et al., 2024), while TC-LIF and TS-LIF introduce coupled dendritic–somatic dynamics (Zhang et al., 2024; Feng et al., 2025). More recent models incorporate synaptic traces to modulate somatic decay, bilinear interactions for neuronal integration, and multi-compartment dendritic branches for sequence modeling (Bai et al., 2026; Ma et al., 2026; Wang et al., 2025a). ConDenFormer builds on the idea of persistent branch-specific memory but introduces *activity- and context-dependent reciprocal conductance exchange*: the soma reads retained dendritic evidence while simultaneously regulating its return flow to the dendrites. This mechanism yields an explicit energy law and enables a source–lag analysis within a native spiking visual backbone.

3 CONDENFORMER

ConDenFormer retains MaxFormer's patch embedding, local token mixing, spiking self-attention, and classifier, while replacing the neuronal computation at its two late transition stages (Figure 1). Each transition neuron consists of a conductance-modulated LIF soma coupled to four persistent dendritic compartments. At each time step, the neuron first exchanges its previously stored somatic and dendritic states, then integrates the current drive and generates a spike, while source-specific evidence is written into the dendrites for the next step (Figure 2).

3.1 SOURCE-PRESERVING INPUTS

We partition the input channels of a native convolution into $R = 4$ disjoint source banks. After the native affine normalization, bank r contributes $p_{r,t}$. We scale each contribution by R to obtain the source write $b_{r,t} = Rp_{r,t}$, while preserving the original net drive

$$x_t = \sum_r p_{r,t} = \frac{1}{R} \sum_r b_{r,t}. \tag{1}$$

In parallel, a companion activity signal $m_{r,t}$ accumulates the absolute synaptic contributions before cancellation. Thus, the model distinguishes a weak net input from strong opposing inputs that happen to produce the same net drive. The interface reuses the native synaptic weights, normalization, and residual connections. At max-pooling locations, each source and activity value follows the location selected by the native net drive. We then apply a bounded, mean-preserving adjustment to obtain the dendritic writes $\widetilde{b}_t$; the implementation is described in Appendix D.1.

3.2 ADAPTIVE EXCHANGE

At step t, let v_t denote the post-reset somatic state and $d_{r,t}$ the dendritic states. Each dendrite has retention ρ_r, capacity C_r, and baseline coupling $g_{0,r}$. Activity and retained context determine its conductance:

$$\begin{aligned} q_{r,t} &= \frac{m_{r,t}}{s_r^m + m_{r,t}}, \qquad z_t = W \,\mathrm{vec}(d_t^{\mathrm{group}}/s^d), \\ g_{r,t} &= g_{0,r}(1 + \alpha_r q_{r,t}) \exp\{\log 4 \tanh z_{r,t}\}, \qquad a_{r,t} = \frac{g_{r,t}}{1 + g_{r,t}}. \end{aligned} \tag{2}$$

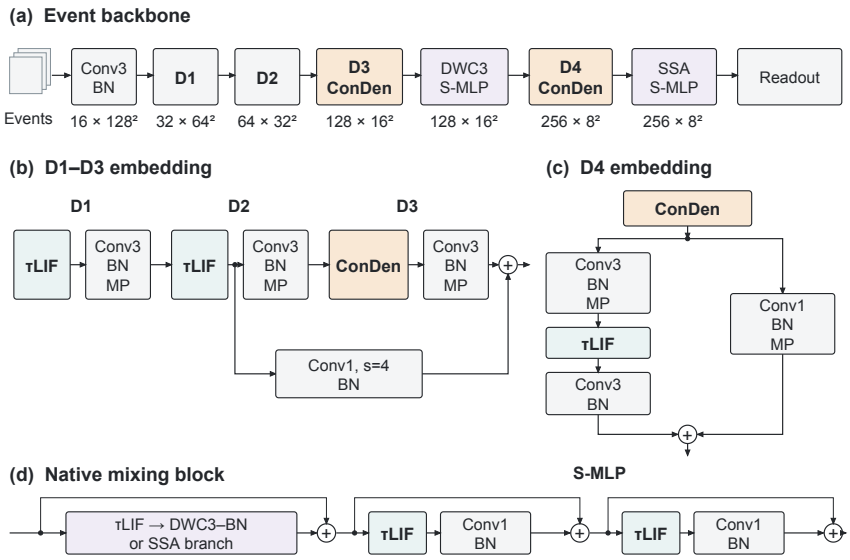

Figure 1: **ConDenFormer architecture.** (a) Event backbone; dimensions denote block outputs. (b) D1/D2 begin with τLIF; D3 begins with ConDen. The D2 shortcut branches after firing and before convolution. (c) Both D4 paths receive the same ConDen spikes; the main path also includes a post-pooling τLIF. (d) Local DWC-3 or global SSA branches retain membrane shortcuts and precede two residual S-MLP projections. ConDen (orange) denotes the conductance LIF soma with persistent dendrites shown in Figure 2; all other neurons, including those in SSA, are learnable-τ LIF. MP: max pooling; Conv$k$: $k \times k$ convolution.

The context group contains four consecutive receiving channels at one location, each with four dendrites. Its 16×16 matrix W is shared across positions. Positive scales s^m, s^d are calibrated from training inputs and frozen; they normalize the gate inputs without normalizing the stored evidence.

The soma and dendrites exchange their states through

$$v_t^m = \frac{v_t + \sum_r a_{r,t} d_{r,t}}{1 + \sum_r a_{r,t}}, \qquad d_{r,t}^m = d_{r,t} + \frac{a_{r,t}}{C_r}(v_t^m - d_{r,t}). \tag{3}$$

With conductance frozen during exchange, this is a closed-form backward-Euler step of a passive star circuit (Appendix A). A dendrite influences the soma and receives a return influence. Increasing its capacity at fixed a_r reduces the change induced by a read.

Current integration, source writing, and firing complete the step:

$$\begin{aligned} u_t &= \beta v_t^m + (1-\beta) x_t, & d_{r,t+1} &= \rho_r d_{r,t}^m + (1-\rho_r)\widetilde{b}_{r,t}, \\ s_t &= \mathbf{1}_{\{u_t \geq \vartheta\}}, & v_{t+1} &= u_t(1-s_t). \end{aligned} \tag{4}$$

Only the soma resets. The present activity controls access to old dendritic states, while newly written evidence becomes available at the next step. History therefore changes the voltage used for firing rather than introducing a separate firing rule.

The effective coefficients of old soma, old dendrites, and current input in u_t are, respectively,

$$\frac{\beta}{1 + \sum_r a_r}, \qquad \frac{\beta a_r}{1 + \sum_j a_j}, \qquad 1 - \beta. \tag{5}$$

Thus, the total conductance shifts the contribution of the exchanged state from the old soma toward retained dendritic evidence, while the relative conductances determine which source histories are preferentially read. The direct contribution of the current input remains fixed at $1 - \beta$. Activity

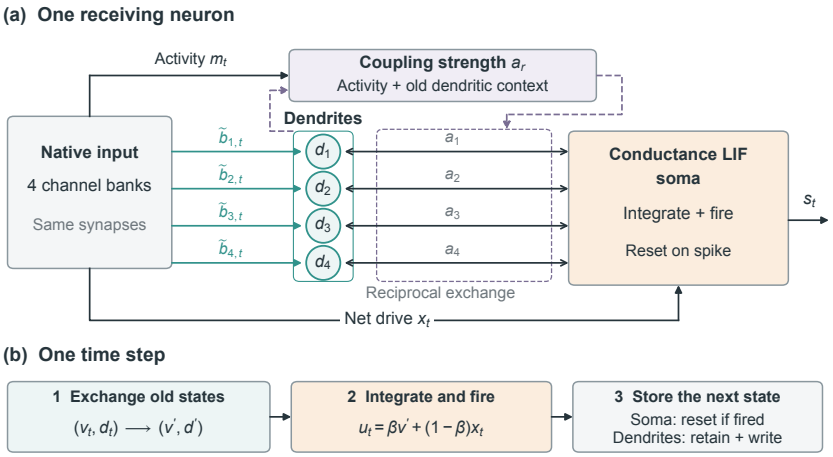

Figure 2: **One conductance LIF soma, four persistent dendrites.** (a) Native synapses provide source writes $\widetilde{b}_{r,t}$, net drive x_t, and pre-cancellation activity m_t. Activity and old dendritic context control all four reciprocal couplings (purple dashed paths). (b) Old states exchange before current-drive integration and firing. Only the soma resets; dendrites retain the exchanged state and receive the current write for the next step. Primes denote post-exchange states.

and dendritic context jointly determine the strength and source preference of the exchange. After a somatic reset ($v_t = 0$), the firing condition at the realized conductances becomes

$$(1-\beta)x_t \geq \vartheta - \frac{\beta \sum_r a_{r,t} d_{r,t}}{1 + \sum_r a_{r,t}}. \tag{6}$$

Positive retained evidence therefore lowers the current drive required to generate a subsequent spike, even when earlier inputs were temporally dispersed. Appendix B illustrates this effect under identical signed inputs and derives the corresponding reduction in the two-step cold-start threshold.

3.3 Implementation

We use Max+/Max patch embedding and DWC-3/SSA token mixing (Fang et al., 2025). The two specialized receiving stages operate at $64\times32\times32$ and $128\times16\times16$ in the event model. Other active neurons use learnable channel-wise retention. A fused FP32 scan performs one update per frame, with mixed-precision convolutions and a sigmoid surrogate of slope four. Appendix D specifies the source interface, parameterization, and native static-image adaptation.

4 Analysis

We first characterize two gates on spike transmission and then show how adaptive exchange enables stable memory access and source–time-selective responses. Proofs and numerical constructions are given in Appendices A–B.

4.1 Timing-sensitive transmission

Consider a current-based LIF receiver with retention $0 < \beta < 1$, threshold $\vartheta > 0$, and zero initial state. Before its first spike, its response to synaptic drives $x_0, \ldots, x_t$ is

$$u_t = (1-\beta)\sum_{k=0}^{t} \beta^{t-k} x_k. \tag{7}$$

Each x_k is the weighted sum of upstream spikes and continuous residual features at that step k. This recurrence gives rise to two distinct limitations that motivate the proposed neuronal dynamics.

Finite-window activation. For a constant nonnegative drive x, the first spike occurs within a T-step simulation window if and only if

$$x \geq \frac{\vartheta}{1-\beta^T}. \tag{8}$$

With hard reset to zero, the exact spike count is $N_T(x) = \lfloor T/n_*(x) \rfloor$, where $n_*(x)$ is the first-crossing time. Figure 3a shows the resulting silent region and its dependence on the observation window. Thus, a spatial feature that is present in the synaptic drive can fail to appear in the emitted spike sequence when its amplitude remains below the finite-window threshold.

Equal counts, different responses. Temporal alignment can also affect spike transmission even when the number of input events is fixed. Suppose two excitatory events each contribute $A < \vartheta$ to the membrane. Coincident events produce a peak of $2A$, whereas separating them by Δ steps produces a second-arrival voltage of $A(1+\beta^{\Delta})$. Hence,

$$A(1+\beta^{\Delta}) < \vartheta \leq 2A \tag{9}$$

yields different spike outputs for input sequences with the same event count. Figure 3b–d illustrates this retention-dependent coincidence window. Conductance-based analyses and cortical experiments likewise identify input synchrony as a determinant of neuronal responses (Becker et al., 2024; Pattadkal et al., 2026). In a visual cascade, variation in first-layer latency can convert spatial structure into such arrival-time differences. Appendix C demonstrates this effect using a two-layer spatial model with matched first-layer spike counts.

Both effects can prevent a receiving neuron from expressing incoming evidence in its output spikes: weak evidence may remain subthreshold throughout the simulation, while temporally dispersed evidence may be attenuated before later inputs arrive. ConDenFormer addresses these two gates by retaining source-resolved subthreshold evidence across somatic spikes and providing adaptive access to that evidence when subsequent inputs arrive.

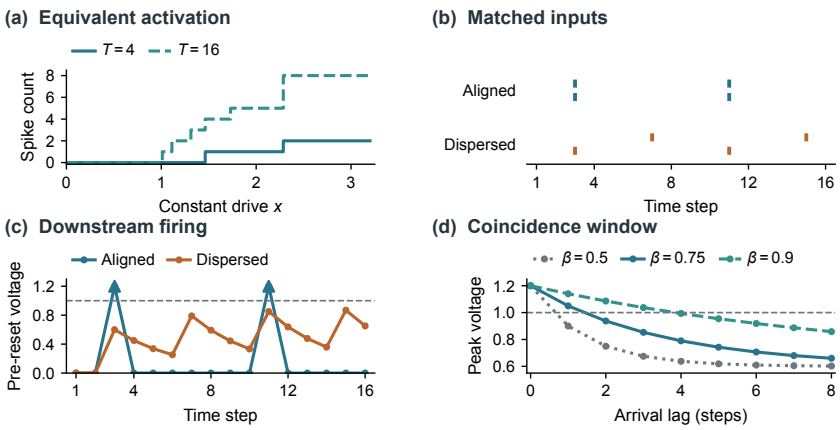

Figure 3: **Two gates on spike transmission.** (a) Exact cold-start spike counts for $T = 4$ and 16, retention $\beta = 0.75$, threshold one, and hard reset. (b) Two sources each provide two identical excitatory events; only arrival times change. (c) Aligned arrivals produce two spikes (triangles), whereas dispersed arrivals produce none. Each event adds $A = 0.6$ to the membrane. (d) The peak $A(1+\beta^{\Delta})$ decreases with arrival lag, defining a retention-dependent coincidence window. Dashed lines in (c,d) mark the firing threshold.

4.2 Dissipative Memory Access

Theorem 1 (Exchange and driven stability). *Let the capacities $C_r \geq 1$ be fixed over a sequence, $0 \leq a_{r,t} < 1$, and $\beta, \rho_r \in (0,1)$. Define charge $Q = v + \sum_r C_r d_r$ and energy $E = \frac{1}{2}v^2 + \frac{1}{2}\sum_r C_r d_r^2$. The exchange in equation 3 preserves Q and satisfies*

$$E(v^m, d^m) - E(v, d) = -\tfrac{1}{2}(v^m - v)^2 - \sum_r a_r \left(1 - \frac{a_r}{2C_r}\right)(v^m - d_r)^2 \leq 0. \tag{10}$$

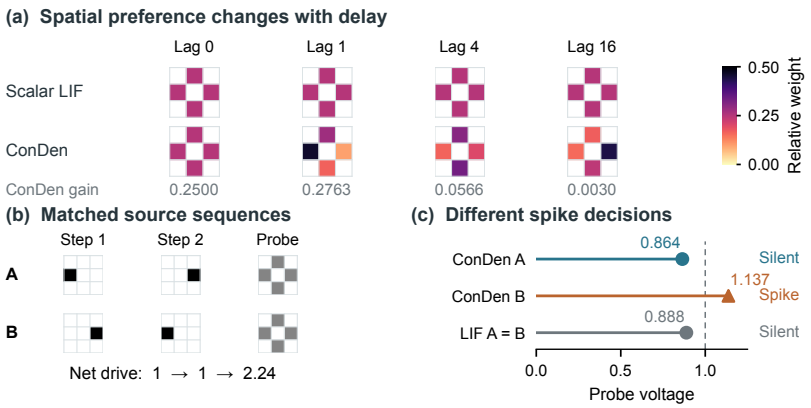

Figure 4: **Source identity and delay jointly determine firing.** (a) Four spatial unit filters visualize the kernel, normalized within each delay on a shared color scale. ConDen's unnormalized gain is reported below; its spatial preference shifts from left to right with delay. (b) Left/right activations of amplitude four are reversed before a common probe of 2.24 per source. Source totals and per-step net drives remain identical. (c) ConDen produces probe voltages 0.864 and 1.137, on opposite sides of threshold one (dashed line); scalar LIF produces 0.888 for both. Triangles denote spikes. Table 4 includes the matched one-way control.

The driven, resetting neuron obeys $E_{t+1} \leq \lambda E_t + U_t$*, where* $\lambda = \max\{\beta, \rho_1, \ldots, \rho_R\} < 1$ *and* U_t *is the weighted energy of current writes. This holds for the adaptive couplings in equation 2.*

The identity gives each operation a distinct role: exchange redistributes evidence, writes supply it, and leakage dissipates it. Bounded writes yield bounded recurrent states. Capacity controls how much redistribution changes the dendritic memory, so access and persistence can be adjusted separately.

4.3 Source–time response

Hold the conductances fixed and consider the subthreshold dynamics. Let $h_t = (v_t, d_{1,t}, \ldots, d_{R,t})^\top$. The recurrence can then be written as

$$h_{t+1} = Ah_t + Bb_t. \tag{11}$$

The contribution of source r to the somatic response at lag ℓ is

$$K(\ell, r) = e_0^\top A^\ell B e_r. \tag{12}$$

A separable kernel has a fixed source preference modulated by a common temporal filter. In contrast, separate dendritic dynamics allow the relative source preference to vary with delay.

Proposition 1 (Nonseparable source–time response)**.** *For* $0 < \beta < 1$*, if* $a_r(1 - \rho_r) \neq a_s(1 - \rho_s)$ *for two sources, then the kernel in equation 12 has rank at least two. A scalar LIF driven by a fixed weighted source sum has kernel rank at most one.*

The condition reflects both writing and reading: dendritic retention determines how evidence enters and persists in memory, while conductance determines how that evidence is subsequently read by the soma. With independent spatial source filters, their composition produces a spatial–temporal response whose spatial preference can change across delays (Figure 4). The adaptive gates further condition this response on incoming activity and retained dendritic context.

As a constructive example, reversing two source activations before a common probe preserves both source totals and per-step net drives, yet produces final probe voltages of 0.864342 and 1.137147, which lie on opposite sides of threshold one. This distinction persists for probe amplitudes $q \in [1.691412, 2.782631)$. Appendix B gives the full construction and a four-state one-way control, separating source-order selectivity from the reciprocal read–write interaction characterized by Theorem 1.

5 EXPERIMENTS

5.1 VISUAL RECOGNITION

Tables 1 and 2 summarize visual recognition results on static images and event streams. Baseline values are taken from the corresponding original papers. For controlled comparisons, we use matched-backbone variants with fixed LIF, learnable-τ LIF, and four-state one-way memory, while keeping the spatial operators, temporal budget, data split, and optimization procedure fixed. Implementation and training details are provided in Appendix D.

ConDenFormer achieves 83.42% on CIFAR-100 with four steps and 85.40% on CIFAR10-DVS with sixteen steps, corresponding to gains of 0.77 and 1.20 percentage points over the reported MaxFormer results, respectively. It also achieves 97.21% on CIFAR-10 and 99.0% on DVS128 Gesture. On ImageNet-1K, ConDenFormer reaches 79.45% with 16.29M parameters and four steps, compared with 77.82% for MaxFormer with 16.23M parameters. These results provide the benchmark context for the controlled studies of memory and adaptive access below.

Table 1: Visual recognition results. Baselines are reported by their original papers. Parameters are in millions; accuracy is Top-1 (%). Spikformer's CIFAR results use its 400-epoch setting.SAFformer uses SFA training with $T \times D = 1 \times 4$ on static-image datasets.

Dataset	Model	Params (M)	Steps	Top-1 (%)
CIFAR-10	Spikformer (2023)	9.32	4	95.51
	SWformer (2024)	7.51	4	96.1
	QKFormer (2024)	6.74	4	96.18
	MSViT (2025)	7.59	4	96.53
	MaxFormer (2025)	6.57	4	97.04
	SpiLiFormer (2025)	7.04	4	96.63
	SAFformer (2026)	6.31	4	97.50
	ConDenFormer (ours)	**6.60**	**4**	**97.21**
CIFAR-100	Spikformer (2023)	9.32	4	78.21
	SWformer (2024)	7.51	4	79.3
	QKFormer (2024)	6.74	4	81.15
	MSViT (2025)	7.59	4	81.98
	MaxFormer (2025)	6.60	4	82.65
	SpiLiFormer (2025)	7.04	4	81.63
	SAFformer (2026)	6.34	4	83.38
	ConDenFormer (ours)	**6.63**	**4**	**83.42**
CIFAR10-DVS	Spikformer (2023)	2.59	16	80.9
	SWformer (2024)	2.05	16	83.9
	QKFormer (2024)	1.50	16	84.0
	MSViT (2025)	1.67	16	84.3
	MaxFormer (2025)	1.45	16	84.2
	SpiLiFormer (2025)	1.57	16	86.7
	SAFformer (2026)	1.50	16	85.0
	ConDenFormer (ours)	**1.47**	**16**	**85.4**
DVS128 Gesture	Spikformer (2023)	2.59	16	98.3
	QKFormer (2024)	1.50	16	98.6
	MSViT (2025)	1.67	16	98.80
	MaxFormer (2025)	1.45	16	98.6
	SAFformer (2026)	1.50	16	99.0
	ConDenFormer (ours)	**1.47**	**16**	**99.0**

5.2 ABLATIONS: MEMORY AND ADAPTIVE ACCESS

We conduct matched-backbone ablations to isolate the effects of membrane dynamics, dendritic memory, and adaptive access (Table 3). All variants use the same spatial operators, membrane shortcuts, and training procedure. Fixed LIF restores conventional LIF neurons, while learnable-τ LIF introduces channel-wise trainable membrane retention without dendritic states. The difference

Table 2: Reported ImageNet-1K classification results at approximately 16M parameters.

Model	Params (M)	Steps	Acc. (%)
Spikformer-8-384 (2023)	16.81	4	70.24
SWformer-6-512 (2024)	18.46	4	74.69
QKFormer (HST-10-384) (2024)	16.47	4	78.80
MSViT-10-384 (2025)	17.69	4	80.09
MaxFormer (Max-10-384) (2025)	16.23	4	77.82
SAFformer-10-384 (2026)	15.12	4	79.19
ConDenFormer (ours)	**16.29**	**4**	**79.45**

between these two controls measures the effect of learnable membrane retention without persistent dendritic memory.

We next examine the role of reciprocal coupling and cross-spike retention. Four-state one-way memory retains the four source-specific memory states and adaptive somatic readout but removes feedback from the soma to the dendrites. Its lower accuracy relative to ConDenFormer isolates the effect of reciprocal exchange while preserving the additional memory states. Spike-reset history retains reciprocal exchange but clears inherited dendritic memory whenever the soma fires, while retaining the current write. Its performance reduction tests whether accumulated evidence remains useful across individual somatic spikes.

Finally, we separate adaptive access from source-specific storage. Static conductance replaces activity- and state-dependent coupling with learned, time-independent conductances while preserving the remaining memory dynamics. The resulting degradation measures the effect of adaptive regulation of access to stored evidence. Shared signed writes provide the same net input to all four dendrites while retaining their separate time constants and adaptive gates. This removes source specificity from the stored values without removing memory or adaptive access. Its performance reduction therefore tests whether maintaining distinct source histories contributes beyond applying multiple temporal filters to a mixed input. Taken together, the ablations distinguish the effects of persistent memory, reciprocal coupling, adaptive access, and source-specific storage.

Table 3: Matched CIFAR10-DVS ablations, $T = 16$. Peak selects the best checkpoint; Last 20 averages epochs 87–106 of the same run.

Neuronal configuration	Peak (%)	Last 20 (%)
Fixed LIF	83.30	82.33
Learnable-τ LIF	83.40	82.69
Four-state one-way memory	82.40	81.30
Spike-reset history	83.40	82.42
Static conductance	82.80	81.35
Shared signed writes	82.60	81.54
ConDenFormer	**85.40**	**83.41**

6 Conclusion

ConDenFormer couples a conductance-based LIF soma with persistent, source-resolved dendritic states. This reciprocal exchange preserves retained evidence and enables source- and time-selective access to it at the soma. Matched-backbone ablations isolate the contributions of reciprocal coupling, cross-spike retention, and adaptive access beyond learnable membrane time constants, while checkpoint interventions connect these neuronal dynamics to firing behavior and correct-class decisions. Together, these results support a neuronal design principle for spiking vision: *preserve evidence at its sources and learn when it should participate in subsequent spikes*.

AI USE STATEMENT

In this work, generative AI tools (specifically ChatGPT 5.6 Sol, ChatGPT 6 Aster, and DeepSeek V4.1 Flash) were used to assist with discovering and summarizing relevant literature, code debugging and optimization, analyzing experimental data (collecting and visualizing metrics and experimental logs in the research process, but not drawing scientific conclusions from the results), and copyediting our final manuscript to improve readability. We also used them to critique the presentation and argumentation of drafts and identify potential weaknesses.

They were not used to propose hypotheses or theoretical models, derive mathematical proofs, design methodology or experiments, assist with translation, clean and reformat datasets, generate synthetic datasets, or interpret results.

AI-generated code and associated experimental results were independently reviewed, tested, and verified by human authors, and we affirm that the code is an accurate implementation of our theory and design. All articles that appear in the reference list were manually checked and carefully read. The human authors take full academic responsibility for all the content of this work.

REPRODUCIBILITY STATEMENT

The appendix specifies the full update, its proofs, numerical constructions, source interface, and training configurations. Meanwhile, we provide anonymized source code in the supplementary material.

REFERENCES

Qianyi Bai, Haiteng Wang, and Qiang Yu. A brain-inspired gating mechanism unlocks robust computation in spiking neural networks. In *International Conference on Learning Representations*, pp. 156097–156129, 2026.

Logan A. Becker, Baowang Li, Nicholas J. Priebe, Eyal Seidemann, and Thibaud Taillefumier. Exact analysis of the subthreshold variability for conductance-based neuronal models with synchronous synaptic inputs. *Physical Review X*, 14(1):011021, 2024.

Tiago Branco, Beverley A. Clark, and Michael Häusser. Dendritic discrimination of temporal input sequences in cortical neurons. *Science*, 329(5999):1671–1675, 2010.

Zhaokun Zhou Han Zhang Jiaqi Wang Huihui Zhou Zhengyu Ma Yonghong Tian Chenlin Zhou, Liutao Yu. Spikingformer: A key foundation model for spiking neural networks. In *International Conference on Learning Representations*, 2026. URL https://openreview.net/forum?id=crsnsaTStJ.

Shikuang Deng, Yuhang Li, Shanghang Zhang, and Shi Gu. Temporal efficient training of spiking neural network via gradient re-weighting. In *International Conference on Learning Representations*, 2022. URL https://openreview.net/forum?id=_XNtisL32jv.

Alain Destexhe, Michael Rudolph, and Denis Paré. The high-conductance state of neocortical neurons in vivo. *Nature Reviews Neuroscience*, 4(9):739–751, 2003.

Wei Fang, Zhaofei Yu, Yanqi Chen, Tiejun Huang, Timothée Masquelier, and Yonghong Tian. Deep residual learning in spiking neural networks. In *Advances in Neural Information Processing Systems*, volume 34, pp. 21056–21069, 2021a.

Wei Fang, Zhaofei Yu, Yanqi Chen, Timothée Masquelier, Tiejun Huang, and Yonghong Tian. Incorporating learnable membrane time constant to enhance learning of spiking neural networks. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 2641–2651, 2021b.

Wei Fang, Yanqi Chen, Jianhao Ding, Zhaofei Yu, Timothée Masquelier, Ding Chen, Liwei Huang, Huihui Zhou, Guoqi Li, and Yonghong Tian. SpikingJelly: An open-source machine learning infrastructure platform for spike-based intelligence. *Science Advances*, 9(40):eadi1480, 2023.

Yuetong Fang, Ziqing Wang, Lingfeng Zhang, Jiahang Cao, Honglei Chen, and Renjing Xu. Spiking wavelet transformer. In *European Conference on Computer Vision*, pp. 19–37, 2024.

Yuetong Fang, Deming Zhou, Ziqing Wang, Hongwei Ren, Zecui Zeng, Lusong Li, Shibo Zhou, and Renjing Xu. Spiking neural networks need high-frequency information. In *Advances in Neural Information Processing Systems*, volume 38, pp. 103404–103425, 2025.

Shibo Feng, Wanjin Feng, Xingyu Gao, Peilin Zhao, and Zhiqi Shen. TS-LIF: A temporal segment spiking neuron network for time series forecasting. In *International Conference on Learning Representations*, pp. 28064–28084, 2025.

Robert Gütig and Haim Sompolinsky. Time-warp–invariant neuronal processing. *PLoS Biology*, 7 (7):e1000141, 2009.

Yifan Hu, Lei Deng, Yujie Wu, Man Yao, and Guoqi Li. Advancing spiking neural networks towards deep residual learning. *IEEE Transactions on Neural Networks and Learning Systems*, 36(2): 2353–2367, 2025.

Wei Hua, Chenlin Zhou, Jibin Wu, Yansong Chua, and Yangyang Shu. MSVIT: Improving spiking vision transformer using multi-scale attention fusion. In *International Joint Conference on Artificial Intelligence*, pp. 5399–5407, 2025.

Hongmin Li, Hanchao Liu, Xiangyang Ji, Guoqi Li, and Luping Shi. CIFAR10-DVS: An event-stream dataset for object classification. *Frontiers in Neuroscience*, 11:309, 2017.

Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. In *International Conference on Learning Representations*, 2019. URL `https://openreview.net/forum?id=Bkg6RiCqY7`.

Jingyang Ma, Chongming Liu, Songting Li, and Douglas Zhou. Beyond linear processing: Dendritic bilinear integration in spiking neural networks. In *International Conference on Learning Representations*, pp. 77297–77323, 2026.

Emre O. Neftci, Hesham Mostafa, and Friedemann Zenke. Surrogate gradient learning in spiking neural networks: Bringing the power of gradient-based optimization to spiking neural networks. *IEEE Signal Processing Magazine*, 36(6):51–63, 2019.

Michalis Pagkalos, Spyridon Chavlis, and Panayiota Poirazi. Introducing the Dendrify framework for incorporating dendrites to spiking neural networks. *Nature Communications*, 14(1):131, 2023.

Jagruti J. Pattadkal, Ronan T. O'Shea, David Hansel, Thibaud Taillefumier, Darrin H. Brager, and Nicholas J. Priebe. Synchrony timescales underlie irregular neocortical spiking. *Neuron*, 114(4): 724–739.e19, 2026. doi: 10.1016/j.neuron.2025.11.005.

Nicolas Perez-Nieves, Vincent C. H. Leung, Pier Luigi Dragotti, and Dan F. M. Goodman. Neural heterogeneity promotes robust learning. *Nature Communications*, 12(1):5791, 2021. doi: 10.1038/s41467-021-26022-3.

Panayiota Poirazi, Terrence Brannon, and Bartlett W. Mel. Pyramidal neuron as two-layer neural network. *Neuron*, 37(6):989–999, 2003.

Alon Polsky, Bartlett W. Mel, and Jackie Schiller. Computational subunits in thin dendrites of pyramidal cells. *Nature Neuroscience*, 7(6):621–627, 2004.

Daniel L. Ruderman. The statistics of natural images. *Network: Computation in Neural Systems*, 5 (4):517–548, 1994.

Balázs B. Ujfalussy, Judit K. Makara, Máté Lengyel, and Tiago Branco. Global and multiplexed dendritic computations under in vivo-like conditions. *Neuron*, 100(3):579–592.e5, 2018.

Kexin Wang, Yuhong Chou, Di Shang, Shijie Mei, Jiahong Zhang, Yanbin Huang, Man Yao, Bo Xu, and Guoqi Li. MMDEND: Dendrite-inspired multi-branch multi-compartment parallel spiking neuron for sequence modeling. In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 27459–27470. Association for Computational Linguistics, 2025a.

Shuai Wang, Malu Zhang, Dehao Zhang, Ammar Belatreche, Yichen Xiao, Yu Liang, Yimeng Shan, Qian Sun, Enqi Zhang, and Yang Yang. Spiking vision transformer with saccadic attention. In *International Conference on Learning Representations*, pp. 24148–24169, 2025b.

Yuchen Wang, Kexin Shi, Chengzhuo Lu, Yuguo Liu, Malu Zhang, and Hong Qu. Spatial-temporal self-attention for asynchronous spiking neural networks. In *Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence*, pp. 3085–3093, 2023.

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.

Zequan Xie, Weiming Zeng, Yunhua Chen, Sichang Lin, Tongyang Chen, and Jinsheng Xiao. SAFformer: Improving spiking transformer via active predictive filtering. In *Proceedings of the Thirty-Fifth International Joint Conference on Artificial Intelligence*, pp. 1902–1910, 2026.

Man Yao, Jiakui Hu, Zhaokun Zhou, Li Yuan, Yonghong Tian, Bo Xu, and Guoqi Li. Spike-driven transformer. In *Advances in Neural Information Processing Systems*, volume 36, pp. 64043–64058, 2023.

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, pp. 16838–16847, 2024.

Hanle Zheng, Yujie Wu, Lei Deng, Yifan Hu, and Guoqi Li. Going deeper with directly-trained larger spiking neural networks. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pp. 11062–11070, 2021. doi: 10.1609/aaai.v35i12.17320.

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.

Zeqi Zheng, Yanchen Huang, Yingchao Yu, Zizheng Zhu, Junfeng Tang, Zhaofei Yu, and Yaochu Jin. SpiLiFormer: Enhancing spiking transformers with lateral inhibition. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 24539–24548, 2025.

Chenlin Zhou, Han Zhang, Zhaokun Zhou, Liutao Yu, Liwei Huang, Xiaopeng Fan, Li Yuan, Zhengyu Ma, Huihui Zhou, and Yonghong Tian. QKFormer: Hierarchical spiking transformer using Q-K attention. In *Advances in Neural Information Processing Systems*, volume 37, pp. 13074–13098, 2024.

Zhaokun Zhou, Yuesheng Zhu, Chao He, Yaowei Wang, Shuicheng Yan, Yonghong Tian, and Li Yuan. Spikformer: When spiking neural network meets transformer. In *International Conference on Learning Representations*, 2023. URL `https://openreview.net/forum?id=frE4fUwz_h`.

A Proofs

A.1 Conductance exchange

Suppress the time index during one exchange and define $\Delta = v^m - v$ and $e_r = v^m - d_r$. The somatic update immediately gives

$$\Delta = -\sum_r a_r e_r, \qquad C_r(d_r^m - d_r) = a_r e_r. \tag{13}$$

Summing these identities proves charge conservation. To obtain the energy law, substitute $v = v^m - \Delta$ and $d_r = v^m - e_r$:

$$\begin{aligned} E^m - E &= v\Delta + \tfrac{1}{2}\Delta^2 + \sum_r \left[a_r d_r e_r + \frac{a_r^2}{2C_r} e_r^2 \right] \\ &= v^m \left(\Delta + \sum_r a_r e_r \right) - \tfrac{1}{2}\Delta^2 - \sum_r a_r \left(1 - \frac{a_r}{2C_r} \right) e_r^2. \end{aligned} \tag{14}$$

The parenthesis vanishes by equation 13; the remaining terms give equation 10. This calculation is pointwise in the conductance, so it also applies when they depend on activity and state.

Circuit interpretation. Reduced compartmental models describe axial communication through voltage-difference currents (Pagkalos et al., 2023). For a unit backward-Euler step of a star circuit,

$$v^m - v = \sum_r G_r(d_r^m - v^m), \qquad C_r(d_r^m - d_r) = G_r(v^m - d_r^m). \tag{15}$$

The second equation yields $d_r^m - d_r = G_r(v^m - d_r)/(C_r + G_r)$. Setting

$$G_r = \frac{C_r a_r}{C_r - a_r} \quad \Longleftrightarrow \quad \frac{G_r}{C_r + G_r} = \frac{a_r}{C_r} \tag{16}$$

recovers equation 3. The closed form implements the circuit exchange without within-frame integration substeps.

Driven bound. For $0 \le t \le 1$, $(tz + (1-t)y)^2 \le tz^2 + (1-t)y^2$. Applying this inequality to the somatic and dendritic writes, then using the energy law, gives

$$E_{t+1} \le \frac{\beta}{2}(v_t^m)^2 + \frac{1}{2}\sum_r C_r \rho_r (d_{r,t}^m)^2 + U_t \le \lambda E_t + U_t,$$
$$U_t = \frac{1-\beta}{2}x_t^2 + \frac{1}{2}\sum_r C_r(1-\rho_r)\widetilde{b}_{r,t}^2. \tag{17}$$

A hard reset can only reduce the somatic squared state. Iteration gives

$$E_t \le \lambda^t E_0 + \sum_{k=0}^{t-1} \lambda^{t-1-k} U_k \le \lambda^t E_0 + \frac{\sup_k U_k}{1-\lambda}, \tag{18}$$

which proves the driven part of Theorem 1.

A.2 Source–time kernel

Let $D = 1 + \sum_r a_r$ and $p^\top = (1, a_1, \ldots, a_R)/D$. Write $M_\gamma = I + \gamma(I - \mathbf{1}\mathbf{1}^\top/R)$ for the implemented mean-preserving source transform. The exact fixed-gate, subthreshold matrices are

$$A_{0,:} = \beta p^\top, \qquad B_{0,:} = \frac{1-\beta}{R}\mathbf{1}^\top,$$
$$A_{r,:} = \rho_r \frac{a_r}{C_r} p^\top + \rho_r \left(1 - \frac{a_r}{C_r}\right) e_r^\top, \qquad B_{r,:} = (1-\rho_r)(M_\gamma)_{r,:}, \quad r = 1, \ldots, R. \tag{19}$$

Here, state index zero denotes the soma. The row sums of A are $\beta, \rho_1, \ldots, \rho_R$, hence $\|A\|_\infty = \lambda < 1$. Starting from zero, the response to a source impulse is exactly $K(\ell, r) = e_0^\top A^\ell B e_r$.

Put $c = (1-\beta)/R$ and $w_r = a_r(1-\rho_r)$. The first two lag rows are

$$K(0, r) = c, \qquad K(1, r) = \frac{\beta}{D}\left[c + (1+\gamma)w_r - \frac{\gamma}{R}\sum_j w_j\right]. \tag{20}$$

Their minor at sources r, s equals

$$\det \begin{pmatrix} K(0,r) & K(0,s) \\ K(1,r) & K(1,s) \end{pmatrix} = \frac{c\beta(1+\gamma)}{D}(w_s - w_r). \tag{21}$$

Since $\gamma \ge 0$, unequal w_r, w_s imply rank at least two. By comparison, scalar LIF driven by $\sum_r \omega_r b_{r,t}$ has kernel $(1-\beta)\beta^\ell \omega_r$, an outer product. This proves Proposition 1.

With affine normalization and pooling selections fixed, spatial source filters $F_r(\delta)$ give $H(\ell, \delta) = \sum_r K(\ell, r) F_r(\delta)$. Linearly independent F_r preserve the source-kernel rank. Figure 4 uses four unit filters at the left, right, upper, and lower positions of a 3×3 neighborhood. These can be realized by four input banks carrying the same image feature with different native spatial weights. The receptive field then changes spatial shape across delays rather than merely rescaling one spatial template. This filter-bank interpretation connects to multiplexed temporal filtering in dendritic models (Ujfalussy et al., 2018); the rank criterion identifies when the filters retain distinct source–delay structure.

B CONSTRUCTIVE NUMERICAL TESTS

B.1 SOURCE ORDER

Use $R = 4$, $\beta = 0.75$, $\rho = (0.1, 0.9, 0.5, 0.75)$, $C_r = 4$, $g_{0,r} = 0.75$, $\alpha = W = 0$, $\gamma = 0.05$, and $\vartheta = 1$. Thus $a_r = 3/7$. Figure 4a shows lags $0, 1, 4, 16$, dividing each kernel row by $\sum_j K(\ell, j)$. Starting from zero, compare

$$A:\ (4,0,0,0),\ (0,4,0,0),\ q\mathbf{1}; \qquad B:\ (0,4,0,0),\ (4,0,0,0),\ q\mathbf{1}. \tag{22}$$

Both sequences have per-step net drive $(1, 1, q)$ and identical per-source totals. Activity and context modulation are disabled to isolate source-specific retention. The first two somatic voltages are $(0.25, 0.756349)$ for A and $(0.25, 0.358454)$ for B. The probe voltages are $u^A = 0.304342 + 0.25q$ and $u^B = 0.577147 + 0.25q$, giving the crossing intervals in Table 4. At $q = 2.24$, the same firing distinction holds for every threshold in $(0.864342, 1.137147]$; the two preceding steps remain subthreshold.

Table 4: Matched source-order test. Threshold is one. Intervals are left-closed and right-open; their endpoints are rounded to six decimals.

Operator	$u^A(q = 2.24)$	$u^B(q = 2.24)$	Order-selective q
Scalar LIF	0.888125	0.888125	$\varnothing$
Four one-way LI	0.849473	1.137423	$[1.690307, 2.842108)$
ConDenFormer	0.864342	1.137147	$[1.691412, 2.782631)$

The four-state control uses the same somatic read but writes $d_{r,t+1} = \rho_r d_{r,t} + (1 - \rho_r)\widetilde{b}_{r,t}$, omitting return influence. Both multistate neurons preserve source order. The comparison isolates source-specific memory; the reciprocal update adds the capacity-dependent storage dynamics analyzed in Theorem 1.

B.2 ACTIVITY-DEPENDENT ACCESS

Equal signed inputs can carry different pre-cancellation activity. With fixed synapses $(1/4, 1, -1)$, inputs $(1, 0, 0)$ and $(1, 1, 1)$ both produce signed contribution $1/4$, but absolute contributions $1/4$ and $9/4$. After the factor $R = 4$, these give the same $b_r = 1$ and different $m_r = 1$ or 9.

For $v = 0$, $d_r = 2$, $x = 1$, $\beta = 0.9$, $g_{0,r} = 0.2$, $\alpha_r = 1$, $s_r^m = 1$, $W = 0$, and $C_r = 4$, the two activities give

m_r	q_r	g_r	a_r	u
1	$1/2$	$3/10$	$3/13$	0.964000
9	$9/10$	$19/50$	$19/69$	1.043448

(23)

The higher activity recruits sufficient stored evidence to cross threshold one. With $\alpha = 0$, both voltages are 0.82. This construction separates activity-mediated access from changes in the current net drive.

B.3 FINITE-WINDOW RECRUITMENT

Start from $v_0 = d_{r,0} = 0$ with constant drive $x > 0$, common source writes $b_r = x$, common retention ρ, and fixed $S = \sum_r a_r > 0$. The contrast write leaves these common source values unchanged. If the first step is silent, $v_1 = (1 - \beta)x$ and $d_{r,1} = (1 - \rho)x$. Substituting into the next exchange gives

$$u_1^{\mathrm{CD}} = \left[1 - \beta^2 + \frac{\beta S}{1 + S}(\beta - \rho)\right] x, \qquad u_1^{\mathrm{LIF}} = (1 - \beta^2)x. \tag{24}$$

Thus, a faster dendritic write ($\rho < \beta$) lowers the two-step first-spike threshold at the same somatic retention:

$$x_{\min}^{\mathrm{CD}}(2) = \frac{\vartheta}{1 - \beta^2 + \beta S(\beta - \rho)/(1 + S)} < \frac{\vartheta}{1 - \beta^2} = x_{\min}^{\mathrm{LIF}}(2). \tag{25}$$

The second-step coefficient exceeds $1 - \beta$, so the first step is indeed silent at this threshold. For $\beta = 0.9$, $\rho = 0.05$, $a_r = 0.5$ for four branches, and $\vartheta = 1$, the thresholds are $10/7$ and $100/19$. At $x = 1.5$, the voltages are respectively 1.05 and 0.285. The read therefore converts evidence into a spike within the shorter window. Capacity affects later steps, when the return influence from this read reaches subsequent decisions.

C TRANSMISSION TESTS

C.1 FINITE-WINDOW ACTIVATION AND EVENT TIMING

For constant x, equation 7 reduces to $u_{n-1} = x(1 - \beta^n)$. Its monotonicity proves equation 8. For $x > \vartheta$, the first-crossing update count and hard-reset spike count are

$$n_*(x) = \left\lceil \frac{\log(1 - \vartheta/x)}{\log \beta} \right\rceil, \qquad N_T(x) = \left\lfloor \frac{T}{n_*(x)} \right\rfloor; \qquad N_T(x) = 0 \text{ for } x \le \vartheta. \tag{26}$$

Figure 3a evaluates this count at 3,201 drives in $[0, 3.2]$, with $\beta = 0.75$, $\vartheta = 1$, and $T \in \{4, 16\}$. The corresponding finite-window drive thresholds are 1.462857 and 1.010124.

In panels b–c, each event adds $A = 0.6$ directly to the membrane. The two sources both fire at steps 3 and 11 in the aligned condition. In the dispersed condition, one keeps steps 3 and 11 while the other uses steps 7 and 15. Retention is 0.75, threshold is one, and the reset is zero. Panel d computes $A(1 + \beta^{\Delta})$ for $\beta \in \{0.5, 0.75, 0.9\}$ and lags $0, \ldots, 8$.

C.2 A SPATIAL CASCADE

Under constant drive, first-spike latency decreases with input strength because

$$\frac{d}{dx} \frac{\log(1 - \vartheta/x)}{\log \beta} = \frac{\vartheta}{x(x - \vartheta) \log \beta} < 0. \tag{27}$$

Local spatial contrasts can therefore become differences in arrival time even when the image is static. We test the resulting cascade on 100×100 periodic gratings with 1–25 cycles across the image, three phases, and two axis-aligned orientations. A normalized 3×3 Gaussian filter ($\sigma = 0.75$) samples each image. Each filtered image is standardized to zero spatial mean and unit spatial standard deviation; this response I gives drive $1.050 + 0.027I$. First-layer LIF neurons use $\beta = 0.75$, threshold one, and 16 input steps. Every first-layer neuron emits exactly one spike.

The second layer receives a positive 9×9 Gaussian-weighted neighborhood ($\sigma = 1.7$, total gain 2.15), with retention 0.5 and two additional tail steps. Its fixed threshold map is sampled from $\mathcal{N}(0.65, 0.07^2)$ and clipped to $[0.48, 0.82]$. For a receiving input sequence E_t, temporal concentration is

$$\mathcal{C} = \frac{\sum_t E_t^2 / (\sum_t E_t)^2 - 1/18}{1 - 1/18}. \tag{28}$$

Figure 5 averages the three phases and two orientations, using the same threshold map for every input. Changes in spatial frequency alter local arrival concentration and second-layer firing despite identical first-layer counts. The controlled example supplies the spatial-to-temporal link used in Section 4.1.

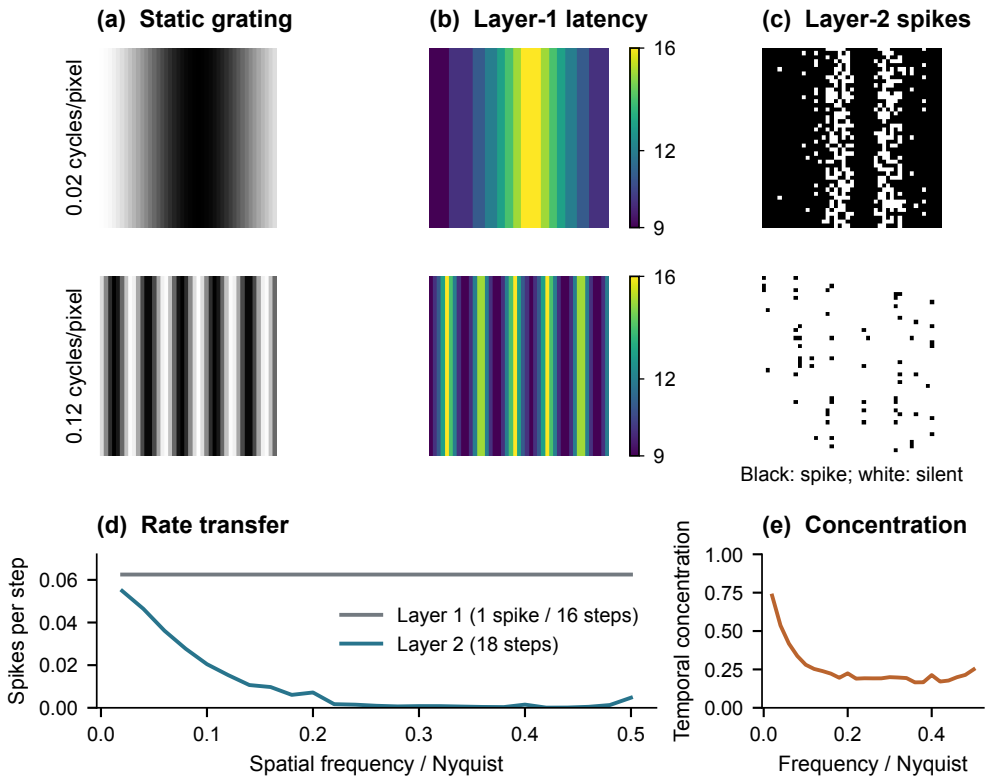

Figure 5: **Spatial structure changes downstream transmission at equal first-layer counts.** Positive local convolution samples static gratings. Spatial frequency changes first-spike timing, the concentration of inputs received by the second layer, and its output firing. All first-layer neurons emit one spike; the downstream difference therefore arises after temporal integration.

D Implementation Details

D.1 Native Source Interface

The source interface carries signed features y and pre-cancellation activity M. On a spike edge, $(y, M) = (s, s)$. The four banks $\mathcal{B}_r$ partition consecutive input channels. For native convolution weights w and bias c, the interface computes

$$p_r = \sum_{j \in \mathcal{B}_r, \delta} w_{j,\delta} y_{j,\delta} + c/R, \quad \mu_r = \sum_{j \in \mathcal{B}_r, \delta} |w_{j,\delta}| M_{j,\delta} + |c|/R, \quad x = \sum_r p_r. \tag{29}$$

Signed and absolute-weight measurements are packed into one grouped convolution. Native BN is evaluated once on x. For its resulting affine map $Ax + c_{\mathrm{BN}}$, each bank receives $p'_r = Ap_r + c_{\mathrm{BN}}/R$ and $\mu'_r = |A|\mu_r + |c_{\mathrm{BN}}|/R$, with a numerical floor $\mu'_r \geq |p'_r|$. Max-pooling selects indices from the net drive once, then gathers every bank at those indices. A continuous residual pair $(y_{\mathrm{res}}, M_{\mathrm{res}})$ adds y_{res} to the net and y_{res}/R, M_{res}/R to each bank's signed and activity terms. The receiver obtains $b_r = Rp_r$ and $m_r = R\mu_r$.

Contrast write. The implemented dendritic write is

$$\widetilde{b} = b + \gamma(b - x\mathbf{1}) = M_\gamma b, \qquad \gamma = 0.25\,\sigma(\eta), \qquad M_\gamma = I + \gamma\left(I - \frac{\mathbf{1}\mathbf{1}^\top}{R}\right). \tag{30}$$

The common source mean is unchanged, while source contrast is scaled by $1 + \gamma$. This operation affects only dendritic writes; current net drive, activity measurements, and pooling selections are unchanged. Each receiving channel has one logit η, initialized to -0.4054651081 so that $\gamma = 0.1$.

D.2 EVENT BACKBONE

Event inputs have shape $[B, 16, 2, 128, 128]$; internal tensors are ordered $[T, B, C, H, W]$. Table 5 lists the complete hierarchy. Pool is 3×3 max-pooling with stride 2 and padding 1. Ordinary active neurons have learnable channel-wise retention.

Table 5: Event architecture. Shapes omit time and batch dimensions.

Stage	Operations	$C \times H \times W$
Stem	Conv3–BN	$16 \times 128 \times 128$
D1	τ-LIF–Conv3–BN–Pool	$32 \times 64 \times 64$
D2	τ-LIF–source Conv3–shared BN–Pool	$64 \times 32 \times 32$
D3	ConDen(64)–Conv3–BN–Pool; add Conv1 stride-4 projection of D2 input spikes	$128 \times 16 \times 16$
Local	DWC3–S-MLP, expansion 1; final MLP projection supplies D4 sources	$128 \times 16 \times 16$
D4	ConDen(128)–Conv3–BN–Pool–τ-LIF–Conv3–BN; add Conv1–BN–Pool of the same D4 spikes	$256 \times 8 \times 8$
Global	SSA–S-MLP, 16 heads, expansion 1	$256 \times 8 \times 8$
Head	Spatial mean–τ-LIF–time mean–Linear	10 or 11 logits

D3 partitions 32 incoming channels into banks of eight; D4 partitions 128 into banks of 32. DWC and S-MLP retain native continuous residuals. SSA uses spiking Q/K/V and $0.125\, Q(K^\top V)$, with attention threshold 0.5 and threshold one elsewhere. Token mixing operates within each frame; neuronal states carry information across frames.

D.3 PARAMETERS AND OPTIMIZATION

Table 6: Special-neuron coordinates at receiving width C_{out}.

Coordinate	Shape	Initialization	Domain
β	C_{out}	0.5	$[0.1, 0.95]$
ρ	$C_{\text{out}} \times 4$	$(0.45, 0.65, 0.8, 0.92)$	$[0.05, 0.97]$
g_0	$C_{\text{out}} \times 4$	0.01	$[0, 1]$
α	$C_{\text{out}} \times 4$	0.5	$[0, 1]$
C	$C_{\text{out}} \times 4$	4	$[1, 8]$
W	$(C_{\text{out}}/4) \times 16 \times 16$	$\mathcal{N}(0, 0.025^2)$	Row $\ell_1 \le 0.5$
γ	C_{out}	0.1	$0.25\,\sigma(\eta)$
s^m, s^d	Each $C_{\text{out}} \times 4$	Calibration	Positive, frozen

All groups read their old dendritic states simultaneously. Direct coordinates are clamped into Table 6's domains after successful optimizer steps. Each row of W is radially rescaled as $W_{i,:} \leftarrow W_{i,:} / \max\{1, \|W_{i,:}\|_1 / 0.5\}$, including after initialization. A dedicated generator with seed $220922 + C_{\text{out}}$ initializes W without advancing the backbone or data random generators. Neuronal coordinates and contrast logits receive no weight decay.

Before optimization, two no-gradient passes first estimate mean branch activity s^m, then the RMS of old dendritic states s^d using the calibrated activity scales. Event calibration uses the same subset of at most 128 training samples in batches of 16, without augmentation, in each pass; its subset seed is the training seed plus 781. Static calibration uses eight training-loader batches per pass, including training transforms. Activity scales at most 10^{-8} and context scales at most 0.05 are replaced by one. BN buffers and random-generator states are restored after calibration, and the scales are frozen.

Spikes use a hard threshold in the forward pass and the derivative $4\sigma(4(u-\vartheta))[1-\sigma(4(u-\vartheta))]$ in backpropagation. Reset is $v_{t+1} = u_t[1 - \text{stopgrad}(s_t)]$. Dendrites persist across spikes; all states reset at the start of each sample. The fused scan uses FP32 state arithmetic under surrounding mixed precision, with one thread per sample, location, and four-channel group.

D3/D4 contain 5,248/10,496 trainable neuronal parameters, adding 15,552 over channel-wise τ-LIF. Dendritic state requires $4(64 \cdot 32^2 + 128 \cdot 16^2) = 393{,}216$ FP32 values per sample (1.5 MiB).

D.4 TRAINING CONFIGURATIONS

The CIFAR and event runs use one NVIDIA GeForce RTX 5090 GPU (32 GB) each, with eight allocated CPUs and 126,000 MB of host memory. Their environment uses Python 3.12.13, PyTorch 2.11.0 with CUDA 13.0, SpikingJelly 0.0.0.0.14, and CuPy 13.6.0 for the fused neuronal scan. The ImageNet-1K protocol uses eight NVIDIA A800 GPUs (80 GB each) interconnected through NVLink, with one distributed-data-parallel process per GPU. TF32 is disabled and neuronal state updates remain in FP32 in both configurations.

Events are integrated by event number into 16 frames at 128×128 (Fang et al., 2023). CIFAR10-DVS (Li et al., 2017) uses the ordered 900/100 examples per class split, Gesture its official 1,176/288 split, and static CIFAR the official 50,000/10,000 split. ImageNet-1K uses its standard training and 50,000-image validation sets. The reference configuration uses seed 35307. All recipes use AdamW (Loshchilov & Hutter, 2019), cosine scheduling, label smoothing 0.1, and batch-mode mixing; Table 7 gives dataset-specific settings.

Table 7: Training configurations. Main epochs include warmup; cooldown is additional. Batch size counts images, not repeated timesteps.

Setting	CIFAR10-DVS	Gesture	CIFAR-100	CIFAR-10	ImageNet-1K
Input resolution	128^2	128^2	32^2	32^2	224^2
Timesteps T	16	16	4	4	4
GPUs	1	1	1	1	8
Batch per GPU	16	16	64	128	64
Global batch	16	16	64	128	512
Peak learning rate	0.006	0.005	0.0015	0.0015	0.0027
Minimum learning rate	10^{-5}	10^{-5}	10^{-5}	10^{-5}	2.5×10^{-6}
Weight decay	0.06	0.06	0.06	0.06	0.05
Main + cooldown epochs	$96 + 10$	$150 + 10$	$400 + 10$	$400 + 10$	$200 + 0$
Warmup epochs	10	10	20	20	5
Initial warmup LR	10^{-5}	10^{-5}	10^{-5}	8×10^{-5}	0
Mixup / CutMix alpha	$0.5/0$	$0.5/0.1$	$0.75/0.5$	$0.5/0.5$	$0.25/1$
Mix / switch probability	$0.5/0.5$	$0.5/0.5$	$1/0.5$	$1/0.5$	$0.5/0.5$
Mixing off (epoch index)	75	75	250	200	Never
Train / eval precision	FP16/FP32	FP16/FP32	FP16/FP16	FP16/FP16	FP16/FP32

Event augmentation uses horizontal flips with probability 0.5 followed by SNN wide augmentation. CIFAR augmentation uses crop/flip, SNN RandAugment `rand-m9-n1-mstd0.4-inc1`, uint8 mixing, normalization, and random erasing with probability 0.25, in this order. A full FP32 CIFAR evaluation is added every ten epochs and at the final epoch. AdamW uses $\epsilon = 10^{-8}$ and betas $(0.9, 0.999)$; event training disables gradient clipping and temporal subsampling.

Eight-GPU optimization. The ImageNet protocol follows the official MaxFormer-10-384 recipe (Fang et al., 2025), with a global batch of $8 \times 64 = 512$ and no gradient accumulation. Its reference learning rate is 0.00135 at batch 256, giving an actual peak rate of $0.00135 \times 512/256 = 0.0027$. A five-epoch linear warmup starts at zero; cosine decay then reaches 2.5×10^{-6} at epoch 200, without additional cooldown. Model weights are trained from scratch; no distillation, exponential moving average, gradient clipping, or stochastic depth is used. Biases, normalization parameters, and neuronal coordinates receive no weight decay. Mixed precision uses FP16 convolutions with dynamic loss scaling and FP32 recurrent states; parameter projection follows each successful optimizer update.

Distributed training partitions images, not their four-step sequences, and resets all neuronal states between minibatches. Native BN statistics are computed locally on each rank, including the folded time and batch dimensions, without SyncBN or separate normalization of the source branches. Before distributed wrapping, rank zero performs the two-pass scale calibration on eight training batches of 64 images per pass. The calibrated activity and context scales and readiness flags are broadcast to every rank; BN buffers and random states are restored. This gives every replica identical frozen neuronal scales. Training samplers reshuffle at every epoch. Validation uses disjoint, non-padded

shards of 6,250 images, aggregating correct counts and sample counts across ranks. Full-precision, single-crop top-1 accuracy selects the checkpoint; no test-time augmentation is used.

D.5 Native static-image model

One augmented RGB image is repeated for $T = 4$; the same crop and mixing transformation are used at every timestep. Neuronal parameters are shared over time, while somatic and dendritic states evolve through the sequence. States begin at zero for each image and are reset before the next minibatch. Static images are processed at their native training resolution, without event-frame conversion or resizing CIFAR inputs to the event-model resolution.

CIFAR. The native 32×32 hierarchy has widths $(96, 192, 384)$, grids $(32^2, 16^2, 8^2)$, and block counts $(1, 1, 2)$: an identity-mixer S-MLP stage, a DWC3 stage, and two SSA blocks with eight heads and MLP expansion four. Special receivers are placed at 96 channels on 32^2 and 192 channels on 16^2, immediately before the two native max-pooling transitions. The classifier remains spatial mean–τ-LIF–time mean–Linear. CIFAR-100 normalization uses mean $(0.5071, 0.4867, 0.4408)$ and standard deviation $(0.2675, 0.2565, 0.2761)$; CIFAR-10 uses $(0.4914, 0.4822, 0.4465)$ and $(0.2470, 0.2435, 0.2616)$.

ImageNet-1K. The extension retains the native MaxFormer-10-384 spatial graph (Fang et al., 2025). A residual convolutional stem reduces 224×224 RGB inputs to $96 \times 56 \times 56$ through two stride-two convolutions. The three stages have widths $(96, 192, 384)$, grids $(56^2, 28^2, 14^2)$, and block counts $(1, 2, 7)$: one DWC7 block, two DWC5 blocks, and seven SSA blocks. All blocks use MLP expansion four, and SSA has six heads of width 64. Two native max-pooling embeddings connect the stages; the continuous membrane shortcuts and the ordering of convolution, BN, pooling, and spiking are unchanged.

The two special receivers replace the input LIF of these embeddings at 96×56^2 and 192×28^2, before downsampling. Four source banks partition the incoming channels of the preceding final MLP projection; shared BN calibration and the continuous residual preserve the native net input. Each receiver maintains four source-specific dendritic states and uses the same bounded coordinates as Table 6. All other active LIF sites, including the classifier neuron, use channel-wise learnable retention. Spatial averaging is followed by the classifier LIF, temporal averaging, and a 1,000-class linear layer. The resulting architecture contains 16,291,720 trainable parameters: 16,237,240 backbone parameters, 30,864 ordinary retention parameters, and 23,616 special-neuron parameters. Temporal replication and distributed replicas do not add trainable parameters.

ImageNet training uses bicubic random-resized crops to 224×224, horizontal flips, RandAugment `rand-m9-mstd0.5-inc1`, and pixel-mode random erasing with probability 0.25. Mixup and CutMix use the settings in Table 7 throughout training. Inputs are normalized with mean $(0.485, 0.456, 0.406)$ and standard deviation $(0.229, 0.224, 0.225)$. Validation resizes the shorter side to 256 and takes a single 224×224 center crop.

E Controlled ablations

Protocol. All component comparisons use the CIFAR10-DVS recipe in Table 7: seed 35307, $T = 16$, batch size 16, peak learning rate 0.006, and 96 main plus 10 cooldown epochs. Ordered sample, label, and split hashes match. The memory/access controls preserve the spatial backbone, continuous shortcuts, shared pooling indices, source calibration procedure, and ordinary learnable-τ neurons. Common backbone weights have identical seeded initialization. Each ablation starts a fresh full training after engineering checks. Table 3 reports this paired seed, separately from the benchmark results in Table 1.

Interventions. The one-way control keeps v_t^m and the adaptive gates unchanged but sets $d_{r,t}^m = d_{r,t}$. Its history update is therefore

$$d_{r,t+1} = \rho_r d_{r,t} + (1 - \rho_r)\widetilde{b}_{r,t}. \tag{31}$$

The now-inactive capacities C_r are frozen. The spike-reset control instead preserves reciprocal exchange and clears only inherited history:

$$d_{r,t+1} = \rho_r[1 - \mathrm{stopgrad}(s_t)]d^m_{r,t} + (1 - \rho_r)\widetilde{b}_{r,t}. \tag{32}$$

Current writes survive the reset. The static-conductance control sets and freezes $\alpha = W = 0$, giving $g_{r,t} = g_{0,r}$; g_0, β, ρ, C remain learnable. It tests joint activity/context adaptation. The shared-write control uses $\widetilde{b}_{r,t} = x_t$ in all four dendrites, retaining distinct retention parameters, reciprocal exchange, and the original source activities for gating. It isolates source-specific signed writes. The fixed-LIF control restores the original $\tau = 2$ neurons throughout the backbone.

Table 8: Run-level ablation details. Epoch indices are one-based. Parameters count trainable coordinates. The CIFAR10-DVS model has 1,474,266 registered parameters, of which 1,474,106 are trainable and 160 are frozen inactive compatibility coordinates.

Configuration	Trainable	Peak epoch	Top-1 (%)
Fixed LIF	1,455,626	95	83.30
Learnable-τ LIF	1,458,554	99	83.40
One-way memory	1,473,338	103	82.40
Spike-reset history	1,474,106	97, 100	83.40
Static conductance	1,461,050	95	82.80
Shared signed writes	1,473,914	96	82.60
ConDenFormer	1,474,106	102	85.40

Trained-state measurements. Evaluation telemetry provides a complementary view of the learned operating points. Over the final 20 epochs, the complete model's mean D3/D4 firing fractions are 0.156/0.208. One-way memory gives 0.145/0.237 and static conductance gives 0.138/0.212, despite their lower accuracies. Spike-reset history learns a lower D4 retention, 0.494 versus 0.758 for the complete model. These measurements are taken on the first evaluation mini-batch; they describe the operating point and do not replace the controlled accuracy comparisons.

F Trained temporal computation

F.1 Protocol

We evaluate the selected fixed-LIF, learnable-τ, and complete-model checkpoints on all 1,000 test examples, reproducing correct predictions. We retain the training implementation, FP32 evaluation, batch size 16, frozen normalization, and RTX 5090 hardware. No weights or normalization statistics are updated. For logits f, the correct-class margin is

$$M = f_y - \max_{j \neq y} f_j. \tag{33}$$

We pair each intervention with the native prediction for the same example. Figure 6 divides margin changes by the native margin standard deviation of the corresponding model. Intervals use 5,000 class-stratified paired bootstrap resamples of the test set.

F.2 From retained evidence to spikes

Let $Z_t = 1 + \sum_r a_{r,t}$. Relative to the same old somatic state and current input without exchange, the exact voltage contribution is

$$\Delta u^{\mathrm{ex}}_t = \frac{\beta}{Z_t} \sum_r a_{r,t}(d_{r,t} - v_t), \qquad \frac{\partial u_t}{\partial a_{r,t}} = \frac{\beta}{Z_t}(d_{r,t} - v^m_t). \tag{34}$$

The partial derivative holds the current states fixed. Exchange can recruit or suppress a spike according to each dendrite's voltage relative to the somatic target. Averaging per-example RMS measurements over the full test set gives Δu^{ex} of 0.322 at D3 and 0.801 at D4, with threshold one. The corresponding local threshold decisions change in 4.37% and 18.28% of neuron–time instances. Native firing fractions are 0.167 and 0.215. These measurements quantify the direct influence of retained evidence on the soma.

F.3 TIMING AND TASK DECISIONS

We circularly shift the binary output of D3 or D4, using either one offset shared by all channels or an independent offset per channel. Both interventions exactly preserve the temporal spike count at every spatial site and channel. Independent shifts alter relative channel phases; a common shift preserves these phases while changing their alignment with the rest of the network. Table 9 gives the resulting accuracies. The stronger effect of independent D3 shifts demonstrates that the trained cascade uses temporal relationships beyond spike counts.

To target dendritic memory, we separately permute the signed writes $\widetilde{b}_{r,t}$ in time or repeat their temporal mean, at both special layers. Current-drive and activity pathways remain intact, and downstream computations are recomputed normally. Shuffled writes change 48 previously correct decisions to errors and correct 23 previous errors, reducing accuracy by 2.50 points. Their mean margin change is -0.126 (95% paired interval $[-0.177, -0.076]$). Mean writes reduce accuracy by 2.70 points. These interventions connect the temporal organization of dendritic evidence to recognition; the independently retrained controls in Table 3 measure its contribution during learning.

Table 9: Frozen-checkpoint timing interventions on the same 1,000 CIFAR10-DVS test examples. Independent and coherent shifts preserve all local spike counts.

Intervention	Fixed LIF	Learnable-τ	ConDenFormer
Native	83.30	83.40	85.40
Input permutation 1	79.80	82.30	83.00
Input permutation 2	81.10	82.20	83.20
D3 coherent shift	81.70	81.80	83.30
D3 independent shifts	76.50	70.50	57.60
D4 coherent shift	83.00	83.20	85.20
D4 independent shifts	78.10	79.70	82.20
Final-spike permutation	83.30	83.40	85.40

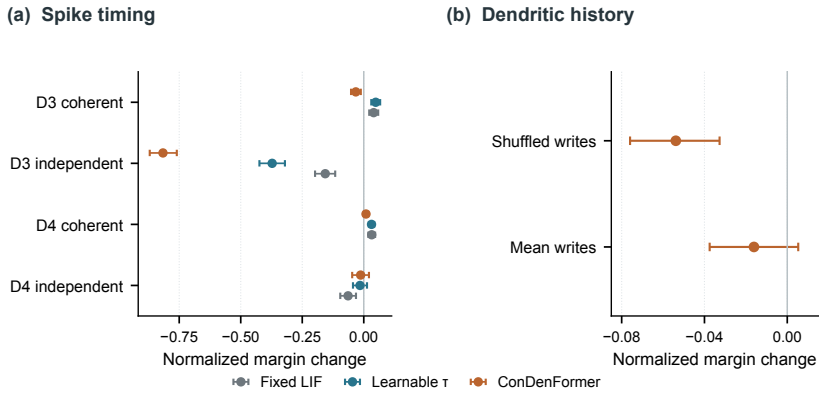

Figure 6: **Temporal computation in trained checkpoints.** (a) Correct-class margin changes after coherent or independent channel shifts. (b) Changes after perturbing only the dendritic-write operator. Points are paired test-set means and bars are 95% bootstrap intervals; changes are normalized within each model. The mean-write interval crosses zero even though accuracy decreases, distinguishing confidence changes from decision changes.

The final classifier is $W(T^{-1}\sum_t s_t^{\mathrm{head}}) + b$. Permuting these final spikes leaves its logits exactly unchanged in all three models. Thus the relevant temporal relationships act before the final average, through state evolution and threshold decisions. Disabling all exchange is a large operating-point perturbation; the static-conductance and one-way retraining controls provide the component comparisons.

F.4 TRAINED SOURCE–LAG RESPONSES

We measure the voltage response to one signed dendritic write along real test trajectories. For each class, we fix the sixth test example before inspecting predictions, and sample every receiving channel at four quarter-grid positions. At step $k = 5$, a central finite difference gives

$$\widehat{K}(\ell, r) = \frac{u_{k+\ell}(\widetilde{b}_{r,k} + \epsilon) - u_{k+\ell}(\widetilde{b}_{r,k} - \epsilon)}{2\epsilon}, \qquad \ell = 1, \ldots, 8. \tag{35}$$

The native update, adaptive gates, and hard resets remain active. We use $\epsilon = 0.1$ and 0.05, retain paths whose four-channel context group has unchanged spike decisions, and require less than 5% relative change between the two estimates. This yields 2,532/2,560 D3 paths and 5,060/5,120 D4 paths. The median relative numerical differences are 4.66×10^{-5} and 4.99×10^{-5}.

For a smooth segment of the trained trajectory, let J_t be its state Jacobian and B_k its signed-write Jacobian. The response contains the ordered product $J_{k+\ell-1} \cdots J_{k+1} B_k$. Source-specific retention and evolving conductance therefore determine how a source's influence changes with delay. We quantify this dependence by

$$\eta = 1 - \frac{\sigma_1(K)^2}{\sum_j \sigma_j(K)^2}, \tag{36}$$

the energy outside the best rank-one source–time kernel. The median η is 0.00762 at D3 and 0.00157 at D4; 40.84% and 9.35% of paths exceed 0.01. Figure 7 shows examples closest to these medians among paths above the first quartile of response norm. Source–lag selectivity is more pronounced at D3, while D4 responses are closer to a shared temporal profile. The task effects are measured separately by the paired interventions above.

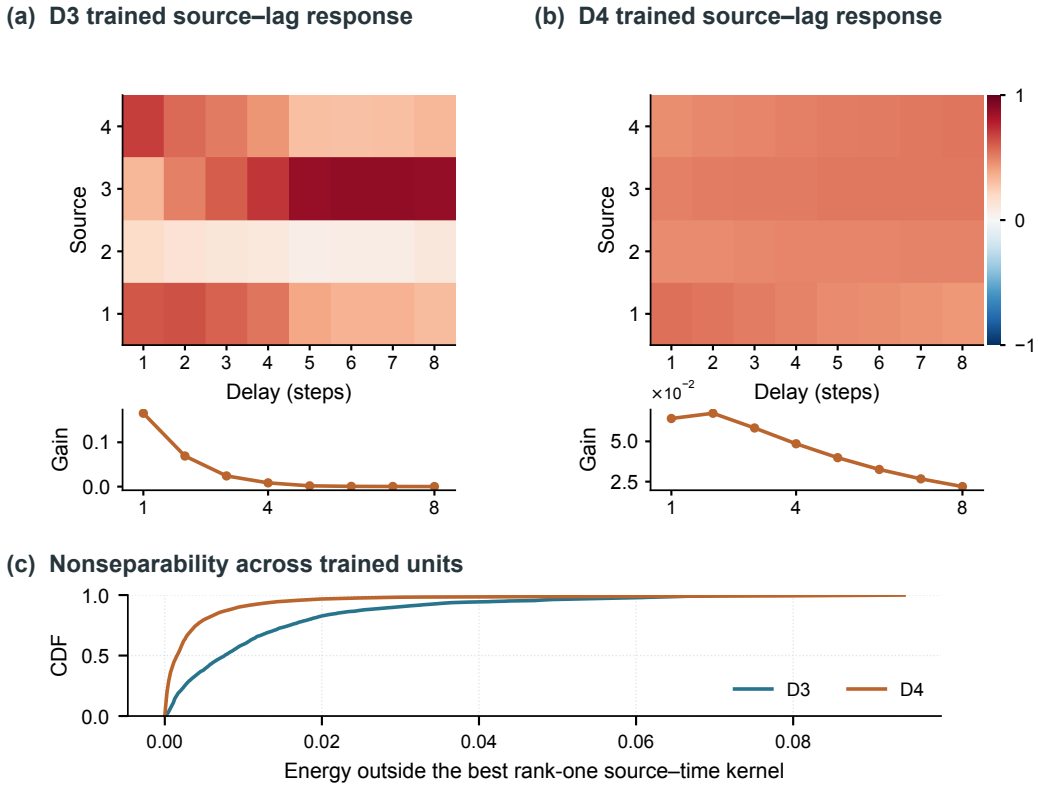

Figure 7: **Source–time responses learned from recognition.** (a,b) D3 and D4 responses to a signed-write perturbation in trained neurons. Heatmaps divide each delay's four-source response by its Euclidean norm, displaying source preference; the curves below retain its absolute gain. (c) Cumulative distribution of nonseparable response energy over all numerically stable sampled paths. The examples follow the median-based selection rule in the text. These are local trained responses, complementing the exact construction in Figure 4.