A SPIKE-DRIVEN HOPFIELD MEMORY FOR ONLINE CONTINUAL LEARNING

Anonymous authors
Paper under double-blind review

ABSTRACT

Associative memory is inspired by the brain's ability to recover stored content from partial cues while incorporating new experiences over time. Hopfield networks provide a computational form of associative recall, yet many variants realise memory through state-based operations rather than spike-mediated neural dynamics. We therefore propose the Spike-driven Hopfield Online Neural Architecture (SHONA), which couples membrane dynamics, timing-dependent plasticity, and spike-driven forgetting control to reformulate Hopfield associative memory as a continual process in which memories are retrieved, updated, and regulated. Experiments show that our approach sustains associative recall while embedding online adaptation directly into the memory process.

1 INTRODUCTION

Associative Memory (AM) is closely related to hippocampal memory function, allowing partial input cues to reactivate stored representations (Rolls, 2013; Horner et al., 2015) for pattern reconstruction. Hopfield (1982) developed Hopfield networks (HNs) as a computational realization of this principle. By encoding memories as attractor states recoverable through recurrent dynamics, HNs provide a natural starting point for biologically inspired memory modelling.

Despite their initial promise, HNs are constrained by binary representations and limited storage capacity. Subsequent models like Dense Associative Memories (DAMs) and Modern Hopfield Networks (MHNs) addressed this capacity bottleneck, with MHNs also moving beyond binary states to continuous representations (Krotov & Hopfield, 2016; Ramsauer et al., 2020). These models can be viewed as generalized retrieval systems, a framework later unified by Universal Hopfield Networks (UHNs) (Millidge et al., 2022), but do not directly address online adaptation.

More recently, Alonso & Krichmar (2024) addressed this limitation with the Sparse Quantized Hopfield Network (SQHN), enabling AM over continual input streams. Yet, its memory operation does not incorporate spike-mediated neural dynamics (Kobayashi & Poo, 2004; Miyata et al., 2013; Wang et al., 2023). Conversely, while the Threshold Phasor Associative Memory (TPAM) demonstrates that neural spike timing can support pattern recovery (Frady & Sommer, 2019), it does not address dynamically storing memories as new inputs arrive. Thus, the integration of spiking mechanism within online AM remains less explored.

To address this gap, we introduce SHONA, a spike-driven online Hopfield architecture that integrates retrieval, memory adaptation, and resilience against catastrophic forgetting (CF) within a unified online AM process.

Our contributions are threefold:

1. We extend Hopfield AM from a static cue-recovery construct into an online memory process that evolves as new inputs arrive.
2. We enable adaptive memory storage through spike-mediated activity and timing-dependent plasticity that directly modify selected memory traces.
3. We introduce a spike-history-driven control mechanism that protects stored traces from destructive modification, mitigating CF during continual updates.

The following sections establish the spike-inspired theoretical foundation underlying SHONA.

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.

2 Related Work

Existing Hopfield-inspired models have evolved to exhibit a broader spectrum of behaviors, yet none support spike-driven, continual memory operation with explicit CF control. The missing intersection between these capabilities motivated the design of SHONA, as highlighted in Figure 1.

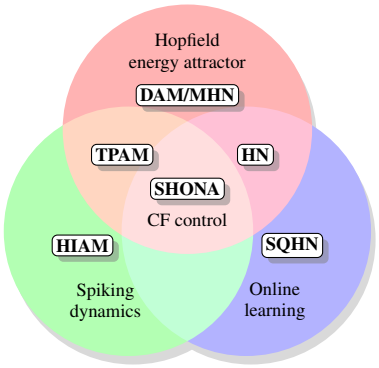


Figure 1: Conceptual positioning of AM models. While existing architectures capture isolated subsets of Hopfield energy attractors, spiking dynamics, and online memory, SHONA combines all three with explicit CF control.

Hopfield (1982) formalized recurrent content-addressable memory by encoding N bipolar patterns $\{\boldsymbol{x}^\mu\}_{\mu=1}^N$, where $\boldsymbol{x}^\mu \in \{-1, 1\}^d$, into a weight matrix $\boldsymbol{W}$:

$$\boldsymbol{W} = \frac{1}{d}\sum_{\mu=1}^{N} \boldsymbol{x}^\mu (\boldsymbol{x}^\mu)^\top, \qquad \boldsymbol{\xi}^{t+1} = \operatorname{sgn}(\boldsymbol{W}\boldsymbol{\xi}^t), \tag{1}$$

driving a corrupted cue $\boldsymbol{\xi}^t$ toward an attractor under the energy function:

$$E(\boldsymbol{\xi}) = -\frac{1}{2}\boldsymbol{\xi}^\top \boldsymbol{W}\boldsymbol{\xi}, \tag{2}$$

This decreases monotonically under asynchronous updates, with capacity $\approx 0.138d$. To overcome this, Krotov & Hopfield (2016) introduced a higher-order energy function in DAM:

$$E(\boldsymbol{\xi}) = -\sum_{\mu=1}^{N} \left((\boldsymbol{x}^\mu)^\top \boldsymbol{\xi}\right)^n, \tag{3}$$

Increasing the polynomial degree $n > 2$ improves capacity, while representations remain bipolar, until Ramsauer et al. (2020) introduced MHN's log-sum-exp energy function with continuous stored patterns $\boldsymbol{X}$ and cue $\boldsymbol{\xi}$. Retrieval becomes:

$$\boldsymbol{\xi}^{\text{new}} = \boldsymbol{X}\,\text{softmax}\left(\beta \boldsymbol{X}^\top \boldsymbol{\xi}\right) \tag{4}$$

where β is an inverse-temperature parameter controlling retrieval sharpness.

Building on this, Millidge et al. (2022) formulated UHN as a general AM retrieval framework, where cue $\boldsymbol{q}$ is compared with stored memories M to obtain similarity scores, which are transformed to emphasize stronger matches before projection into the retrieved output.

$$\boldsymbol{z} = P(\text{sep}(\text{sim}(M, \boldsymbol{q}))) \tag{5}$$

However, UHN takes M as given and does not specify subsequent updates, leaving continual memory adaptation outside the framework. Alonso & Krichmar (2024) addressed this in SQHN by incrementally updating the connection matrix $M_{pa_l,l}$ at each training iteration:

$$\Delta M_{pa_l,l} = \frac{1}{c^*_{pa_l}+1}\left(h^*_l - M_{pa_l,l}h^*_{pa_l}\right) h^{*\top}_{pa_l} \tag{6}$$

where h^*_l and $h^*_{pa_l}$ are the current node assignments and $c^*_{pa_l}$ counts prior parent activations, yielding a local continual update as new samples arrive.

This approach, however, relies on shifted-cosine matching, probabilistic energy maximisation, and capacity expansion through neuron growth. Consequently, it departs from classical Hopfield attractor dynamics and does not integrate spiking dynamics into memory operation. Conversely, Frady & Sommer (2019) incorporates spiking dynamics through TPAM while retaining fixed-point attractor dynamics governed by a Hopfield-like energy function. By representing neural states as complex phasors $z_i(t)$, TPAM maps their phases to explicit spike times $s_i^{(t)}$:

$$E(\boldsymbol{z}) = -\frac{1}{2}\sum_{ij} W_{ij} z_i z_j^* + \Theta\|\boldsymbol{z}\|_1, \qquad s_i^{(t)} = T\left(\frac{z_i^{\phi}(t)}{2\pi} + t\right) \tag{7}$$

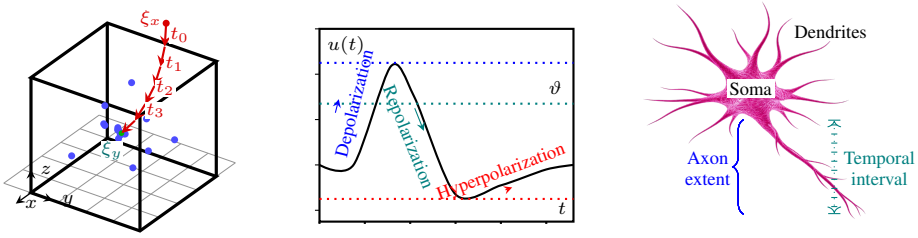
However, TPAM remains a fixed-memory retrieval model, with associative weights that do not adapt as new inputs arrive. A comparable limitation appears in hippocampus-inspired associative memory (HIAM) (Casanueva-Morato et al., 2024). While HIAM uses Spike-Timing-Dependent Plasticity (STDP) to form cue–content associations, it does not explicitly preserve previously stored memories, allowing existing memories to be overwritten as new ones are learned. This undermines continual learning and, like TPAM, fails to provide a mechanism for continual spike-driven AM.

3 Preliminaries

In Hopfield-style AM, retrieval is an iterative process where an input cue gradually converges toward a stable attractor state (Hopfield, 1982; Ramsauer et al., 2020), as illustrated in Figure 2a. Alternatively, neuronal membrane polarization in Figure 2b offer a view for mapping an incoming cue to stored memory traces. This process can be mathematically abstracted by the Adaptive Leaky Integrate-and-Fire (ALIF) model (Bellec et al., 2018; Dayan & Abbott, 2005) as follows:

$$u(t+1) = \alpha u(t)\big(1 - s(t)\big) + Wx(t), \qquad \vartheta(t+1) = \beta\vartheta(t) + \delta\vartheta\, s(t) \tag{8}$$

Here, the membrane potential $u(t)$ is evaluated against the adaptive threshold $\vartheta(t)$ to produce the binary spike output $s(t) \in 0, 1$, yielding a time-dependent neural response.



(a) Hopfield retrieval dynamics **(b)** Neuronal polarization **(c)** Axonal extent and temporal interval

Figure 2: (a) Hopfield retrieval as iterative state evolution from cue ξ_x toward stored attractor ξ_y. (b) Neuronal polarization as time-dependent membrane evolution through depolarization, firing, repolarization, and hyperpolarization. Together, (a) and (b) illustrate complementary views of memory retrieval through global attractor convergence and temporal neural response. (c) Extending this temporal view to spike propagation, axonal extent influences propagation time (Debanne et al., 2011) and can itself vary structurally (Luo & O'Leary, 2005), motivating a variable temporal span.

While neuronal polarization provides a basis for cue–content association, STDP enables the AM trace to be modified according to the relative timing of pre- and post-synaptic spikes (Markram et al., 1997; Bi & Poo, 1998):

$$\Delta W_{ij} = \begin{cases} A_+ \exp\left(-\frac{\Delta t}{\tau_+}\right), & \Delta t > 0, \\ -A_- \exp\left(\frac{\Delta t}{\tau_-}\right), & \Delta t < 0, \end{cases} \tag{9}$$

where Δt captures this timing difference, through which the AM trace can be gradually updated.

Given that synaptic modification depends on spike events distributed over time (Bi & Rubin, 2005; Dan & Poo, 2006), the temporal extent over which these interactions occur becomes relevant. Biological axons can themselves undergo structural remodelling through elongation and retraction (Luo & O'Leary, 2005; De Paola et al., 2006), motivating a temporal extent that is variable rather than

fixed, as conceptually illustrated in Figure 2c. Such variation can be regulated through the bounded nonlinear response of Hill (1910); Naka & Rushton (1966):

$$\frac{R(x)}{R_{\max}} = \frac{x^n}{(x^n + K^n)} \tag{10}$$

where K determines the response scale and n its steepness. The normalized response is bounded between zero and one, providing a controlled measure for adjusting the temporal extent.

4 METHODOLOGIES

4.1 SPIKE-DRIVEN SIMILARITY

Associative retrieval is built on how stored memory responds to an incoming cue. SHONA realizes this cue–memory as a similarity mapping through biologically inspired spiking dynamics. Inspired by LIF, each input cue is encoded as a spike train $\boldsymbol{x}(t) \in \{0,1\}^d$ over a time horizon T via rate coding (Appendix A.1), where its similarity to each stored trace $\boldsymbol{\xi}_i$ evolves through the membrane potential $u_i(t)$, while an adaptive threshold $\vartheta_i(t)$ regulates firing for winner selection.

$$u_i(t) = u_i(t-1) + \boldsymbol{\xi}_i^\top \boldsymbol{x}(t) - \eta_u \mathbb{I}\left[u_i(t-1) + \boldsymbol{\xi}_i^\top \boldsymbol{x}(t) > u_{\text{rest}}\right] \qquad u_i(0) = u_{\text{rest}}, \tag{11a}$$

$$\vartheta_i(t) = \max\left\{\vartheta,\, \vartheta_i(t-1) - \eta_\vartheta \mathbb{I}[u_i(t) > u_{\text{rest}}]\right\} \qquad \vartheta_i(0) = \vartheta. \tag{11b}$$

Here, i indexes stored traces and $t = 1, \ldots, T$ denotes spike time steps, while η_u and η_ϑ control membrane leakage and threshold dynamics, respectively, governing how accumulated similarity evolves.

At time t, all traces whose membrane potentials exceed their adaptive thresholds enter the active set $A(t)$; if multiple traces are active, the trace with the strongest membrane response is selected as the winner:

$$A(t) = \{i : u_i(t) > \vartheta_i(t)\}, \qquad i^*(t) = \arg\max_{i \in A(t)} u_i(t). \tag{12}$$

where, $i^*(t)$ indexes the winning trace at t; if $A(t) = \varnothing$, accumulated similarity is insufficient to select one.

Once selected, the winner is suppressed by hyperpolarisation and threshold elevation, while lateral inhibition regulates competing responses:

$$u_i(t^+) = \begin{cases} u_{\text{hyp}}, & i = i^*(t), \\ u_{\text{inh}}, & i \neq i^*(t), \end{cases} \qquad \vartheta_{i^*(t)}(t^+) = \vartheta_{i^*(t)}(t) + \delta_\vartheta. \tag{13}$$

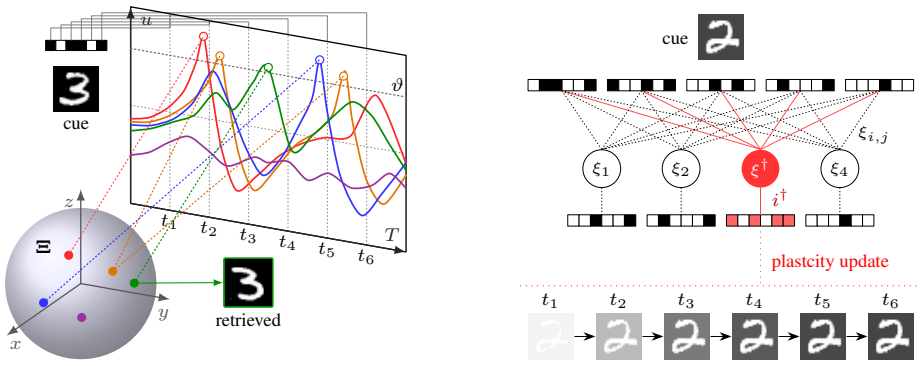
Specifically, u_{hyp} and u_{inh} set the post-firing levels of winning and competing traces, while δ_ϑ raises the winner threshold and τ_{ref} sets its refractory duration to prevent immediate refiring. SHONA, through Equations 11–13, demonstrates that cue–trace similarity can be realized by membrane polarisation and winner firing, as illustrated in Figure 4a, Despite replacing conventional basin-attractor dynamics with spike-driven competition, the resulting retrieval still follows Hopfield-style energy descent while retaining exponential storage-capacity scaling, as established in Appendices A.2 and A.3, respectively.

4.2 WINNER-CONDITIONED MEMORY UPDATE

As an online associative memory, SHONA continuously adapts its stored traces through an STDP-inspired causal rule, whereby the trace that best matches an incoming cue under the preceding similarity mechanism is selected and progressively reshaped toward that cue through the following winner-conditioned plasticity rule:

$$\xi_{i^*(t),k} \leftarrow \xi_{i^*(t),k} + \sigma \begin{cases} \left(\xi_{\max} - \xi_{i^*(t),k}\right)^\mu A_- \left(\exp\left(\frac{\Delta t_k(t)}{\tau_-}\right) - \rho\right), & \Delta t_k(t) \leq 0, \\ -\left(\xi_{i^*(t),k} - \xi_{\min}\right)^\mu A_+ \left(\exp\left(-\frac{\Delta t_k(t)}{\tau_+}\right) - \rho\right), & \Delta t_k(t) > 0. \end{cases} \tag{14}$$

where $\xi_{\max}$ and $\xi_{\min}$ bound trace strengthening and weakening by progressively reducing plasticity as either limit is approached, consistent with bounded STDP formulations (Gütig et al., 2003;



(a) SHONA spiking retrieval dynamics **(b)** Winner-conditioned plasticity update

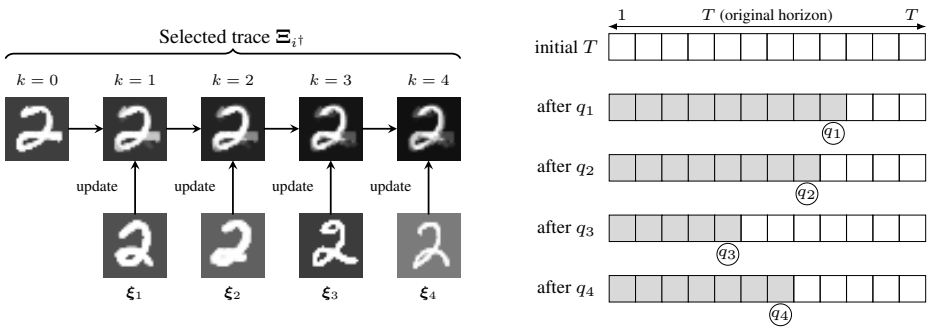
Figure 3: SHONA spike-driven retrieval and winner-conditioned plasticity. (a) Rate-coded input spikes accumulate cue–trace similarity in trace-specific membrane responses, with adaptive thresholds governing firing and the strongest active trace selected as the winner. (b) The encoded cue is compared across stored traces ξ_i, with the winning trace $\xi_{i^\dagger}$ determining the active update pathway. Its synaptic representation is then modified through STDP plasticity over successive time steps $t_1, \ldots, t_6$, progressively adapting the stored trace toward the current cue.

Van Rossum et al., 2000). The spike timing used in the update is obtained from the most recent presynaptic spike over the preceding P time steps:

$$\Delta t_k(t) = \begin{cases} \tau_k^*(t) - t, & \text{if a spike occurs in } [t-P, t], \\ 1, & \text{otherwise.} \end{cases} \tag{15a}$$

$$\tau_k^*(t) = \max\left\{ s \in [t-P, t] : x_k(s) = 1 \right\}, \tag{15b}$$

Here, $\tau_k^*(t)$ identifies the most recent presynaptic spike used for the update. If such a spike is found, $\Delta t_k(t) \leq 0$ increases the corresponding value of the winning trace; otherwise, $\Delta t_k(t) = 1$ decreases it. The trace is thus progressively shaped by recurring cue activity. Retaining only the most recent spike yields a simplified nearest-neighbour STDP rule (Morrison et al., 2008). The winner-conditioned update is illustrated in Figure 4a, with its semantics summarised in Figure 3b.



(a) SHONA winner-conditioned update **(b)** SHONA Adaptive catastrophic horizon

Figure 4: SHONA update and adaptive horizon control. (a) Successive cues $\boldsymbol{\xi}_k$ update the same selected trace $\boldsymbol{\Xi}_{i^\dagger}$ through winner-conditioned plasticity, with k denoting its successive update states. (b) After each winner-firing event q_k, the effective horizon $T_{\text{new}}^{(k)}$ is recomputed from the inter-spike interval, winner count c, and accumulated usage a, truncating the remaining update window. Gray, highlighted, and blank boxes denote elapsed, retained, and truncated steps, respectively.

4.3 Catastrophic Forgetting Control

A consequence of continual plasticity is that repeated updates can gradually overwrite stored traces, leading to CF. SHONA mitigates this drift by limiting the duration over which a repeatedly selected trace remains available for further plasticity. To regulate this interval, SHONA tracks three winner-dependent signals during online updating: the interval between successive winner events ι_k, the selection count $c_{i,k}$ of trace i, and its accumulated usage a_i, with detailed definitions provided in

Appendix A.4.

$$\iota_k = t_k - t_{k-1}, \qquad c_{i,k} = \sum_{r=1}^{k} \mathbb{I}[i_r = i], \qquad a_i \leftarrow a_i + \frac{c_i}{(\max_j c_j + \epsilon)}. \tag{16}$$

Here, c_i denotes the final value of $c_{i,k}$ for the current input. Together, these signals capture winner-event spacing, repeated selection of trace i, and its longer-term usage, providing the basis for determining how much of the original horizon T is retained for subsequent updating:

$$T_{\text{new}} = \lfloor T \times g(c_{i,k}) \times g(\iota_k) \times g(a_i) \rfloor \leq T \tag{17}$$

Here, $g(\cdot)$ is the bounded nonlinear response following Hill (1910); Naka & Rushton (1966), as introduced in Section 3:

$$g(x) = b + (1-b)\left(\frac{x^n}{x^n + K^n}\right). \tag{18}$$

where K and n follow Equation 10, and b sets the lower bound of the response. Since $g(x) \in [b, 1]$, the effective horizon remains bounded by T, so that $T_{\text{new}} \leq T$. Recomputing T_{new} after each winner event therefore restricts further plasticity for repeatedly selected traces, reducing progressive memory overwrite. If left unchecked, persistent trace dominance can bias subsequent retrieval and concentrate further plasticity on the same trace. As a secondary safeguard, SHONA identifies the most frequently selected trace and resets it when either its accumulated usage or consecutive-selection streak exceeds the corresponding threshold:

$$i^\dagger = \arg\max_i c_i, \qquad a_{i^\dagger} > \tau_a \;\text{ or }\; s_{i^\dagger} > \tau_s. \tag{19}$$

Here, $a_{i^\dagger}$ tracks the accumulated usage of the most frequently selected trace, while $s_{i^\dagger}$ measures its persistence across consecutive inputs; τ_a and τ_s define their respective reset thresholds. Exceeding either threshold resets the trace, preventing repeated selection from biasing subsequent retrieval and further modifying the same trace. The adaptive-horizon process is illustrated in Figure 4b, with the corresponding dominance-pruning bound established in Appendix A.5, while the complete SHONA pseudocode is provided in Appendix A.9.

5 Experiments

5.1 Experimental Setup

We examine SHONA's ability to construct and retain memories across grayscale, RGB, and event-based data, evaluating both retrieval from the evolving memory and preservation of previously stored information. Experiments use MNIST, EMNIST, Fashion-MNIST (FMNIST), MNIST+Fashion-MNIST (F+MNIST), CIFAR-10, CIFAR-100, and CIFAR10-DVS at three memory sizes: $300/1300/2300$ for grayscale, $700/2000/3300$ for RGB, and $150/300/500$ for DVS inputs. Retrieval is evaluated using clean cues, Gaussian noise, and masking, with the primary evaluation using $\sigma = 0.5$ noise and 50% masking. The default readout is First-Firing Winner (FFW), with alternative temporal readouts detailed in Appendix A.6. Comparisons are made against MHN, SQHN, and cosine similarity. Performance is measured using MSE and SSIM; CF ablations compare Full-CF, No-CF, Horizon-only, and Pruning-only

5.2 Associative Retrieval Performance

Table 1 reports online retrieval performance using FFW. Across MNIST, EMNIST, FMNIST, and F+MNIST at sizes of 1300 and 2300, SHONA outperforms MHN in all 24 clean, noisy, and masked evaluations. At 2300, clean-cue MSE is 57.0–65.3% lower than MHN, while under 50% masking it is 42.9% and 53.2% lower on F+MNIST and EMNIST, respectively.

The strongest relative results against all three baselines occur under incomplete cues. Against the best of MHN, SQHN, and cosine, SHONA achieves the lowest MSE in six masked settings, including all three FMNIST sizes, with improvements reaching 21.0%. SHONA also achieves the lowest noisy-cue MSE on CIFAR10-DVS at size 150.

Table 1: Online retrieval MSE using FFW readout. Each entry reports on *Clean/Noise/Mask*. Noise uses Gaussian noise with $\sigma = 0.5$, and Mask removes 50% of the cue.

Dataset	Size	SHONA	MHN	SQHN	Cosine
MNIST	300	0.1389 / 0.1090 / 0.1459	0.0999 / 0.1052 / 0.1010	0.1272 / 0.1298 / 0.1528	0.0976 / 0.1016 / 0.1122
	1300	0.0570 / 0.1043 / 0.0631	0.1099 / 0.1121 / 0.1086	0.0370 / 0.0398 / 0.1204	0.0307 / 0.0494 / 0.0476
	2300	0.0480 / 0.1239 / 0.0625	0.1267 / 0.1280 / 0.1253	0.0302 / 0.0345 / 0.1381	0.0279 / 0.0512 / 0.0460
EMNIST	300	0.1414 / 0.1237 / 0.1509	0.1186 / 0.1222 / 0.1322	0.1295 / 0.1305 / 0.1600	0.1108 / 0.1190 / 0.1410
	1300	0.0618 / 0.1037 / 0.0812	0.1140 / 0.1287 / 0.1140	0.0446 / 0.0485 / 0.1257	0.0428 / 0.0585 / 0.0921
	2300	0.0663 / 0.1780 / 0.0859	0.1909 / 0.2027 / 0.1837	0.0393 / 0.0423 / 0.1290	0.0388 / 0.0658 / 0.0931
FMNIST	300	0.1453 / 0.1125 / 0.1681	0.1112 / 0.1150 / 0.1833	0.0933 / 0.0946 / 0.1888	0.0940 / 0.0992 / 0.1868
	1300	0.0882 / 0.1644 / 0.1120	0.1803 / 0.1991 / 0.1588	0.0333 / 0.0362 / 0.1418	0.0327 / 0.0504 / 0.1496
	2300	0.0937 / 0.2220 / 0.1203	0.2368 / 0.2492 / 0.2046	0.0307 / 0.0335 / 0.1370	0.0301 / 0.0499 / 0.1375
F+MNIST	300	0.1571 / 0.1381 / 0.1619	0.1363 / 0.1405 / 0.1502	0.1235 / 0.1258 / 0.1678	0.1127 / 0.1222 / 0.1556
	1300	0.0773 / 0.1446 / 0.1168	0.1579 / 0.1670 / 0.1740	0.0403 / 0.0431 / 0.1135	0.0389 / 0.0596 / 0.1393
	2300	0.1143 / 0.2612 / 0.1177	0.2658 / 0.2697 / 0.2062	0.0343 / 0.0383 / 0.1208	0.0345 / 0.0618 / 0.1574
CIFAR10	700	0.1237 / 0.0540 / 0.1292	0.0526 / 0.0526 / 0.0974	0.0529 / 0.0538 / 0.1537	0.0555 / 0.0553 / 0.1375
	2000	0.0385 / 0.0174 / 0.0405	0.0168 / 0.0169 / 0.0249	0.0132 / 0.0141 / 0.1527	0.0133 / 0.0139 / 0.1339
	3300	0.0282 / 0.0186 / 0.0267	0.0184 / 0.0184 / 0.0220	0.0119 / 0.0129 / 0.0899	0.0120 / 0.0132 / 0.0873
CIFAR100	700	0.1278 / 0.0589 / 0.1342	0.0579 / 0.0580 / 0.1303	0.0571 / 0.0582 / 0.1928	0.0592 / 0.0594 / 0.1463
	2000	0.0387 / 0.0186 / 0.0420	0.0182 / 0.0183 / 0.0305	0.0136 / 0.0143 / 0.1338	0.0139 / 0.0143 / 0.1276
	3300	0.0308 / 0.0203 / 0.0324	0.0202 / 0.0202 / 0.0233	0.0121 / 0.0132 / 0.1262	0.0123 / 0.0139 / 0.1264
CIFARDVS	150	0.2722 / 0.2523 / 0.2814	0.2521 / 0.2531 / 0.2723	0.2697 / 0.2698 / 0.2943	0.2570 / 0.2563 / 0.2891
	300	0.1613 / 0.1181 / 0.1627	0.1183 / 0.1211 / 0.1361	0.1221 / 0.1220 / 0.1556	0.1164 / 0.1147 / 0.1873
	500	0.1667 / 0.1312 / 0.1618	0.1324 / 0.1358 / 0.1320	0.1072 / 0.1070 / 0.1734	0.1013 / 0.1049 / 0.1707

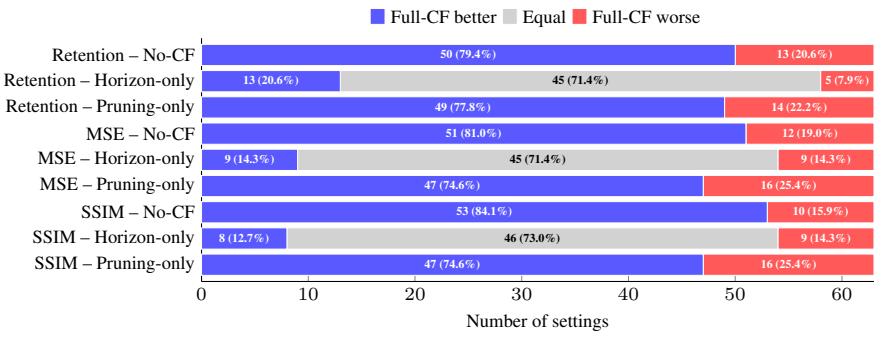
The same pattern is reflected in the SSIM results reported in Table 4 in Appendix A.7.1. SHONA achieves the highest SSIM among the four methods in $11/21$ masked evaluations and exceeds MHN in $41/63$ evaluations overall, showing that its performance under incomplete cues is not limited to pixel-wise error.

Overall, SQHN and cosine provide strong similarity-based comparisons, while MHN provides the most directly aligned baseline for SHONA's Hopfield-style retrieval. Against MHN, SHONA demonstrates strong retrieval at larger grayscale memory sizes and under masking while retaining its spike-driven online dynamics.

5.3 Continual Memory Retention and Forgetting Control

We evaluate retention of the same early set of 50 stored memories during continued online updates, measuring how well previously acquired information is preserved. Figure 5 shows that Full-CF achieves lower retention MSE than No-CF in $50/63$ settings and than Pruning-only in $49/63$. Detailed per-dataset retention results and checkpoint trajectories are provided in Table 6 and Figure 8 in Appendix A.7.2.

Figure 5: Comparison of Full-CF against the three CF ablations across 63 dataset, memory-size, and cue settings. Lower MSE and higher SSIM are better. Horizon-only denotes horizon enabled without pruning, while Pruning-only denotes pruning enabled without the horizon.



The ablations distinguish the roles of the two CF mechanisms. Horizon-only matches Full-CF in $45/63$ retention settings, whereas Full-CF outperforms Pruning-only in $49/63$, indicating that the

adaptive horizon provides the primary protection against progressive overwrite. Pruning therefore acts as a complementary safeguard against persistent trace dominance.

Beyond retention, Full-CF outperforms No-CF in $51/63$ online MSE and $53/63$ SSIM evaluations, showing that CF control also benefits online retrieval in most settings.

5.4 Temporal Retrieval and Readout Ablation

We next examine how temporal readout affects retrieval from SHONA's spike-driven dynamics. We consider four readout strategies, detailed in Appendix A.6. Table 2 summarises the number of settings in which each strategy achieves the best or tied-best MSE and SSIM. Overall, FFW performs best most frequently, attaining the best or tied-best MSE in $35/63$ settings and SSIM in $45/63$, supporting its use as the default readout in the preceding experiments. Additional results are reported in Appendix A.7.3.

Table 2: Readout ablation under online retrieval with Full-CF. Entries report the number of settings in which each readout obtains the best or tied-best MSE/SSIM.

Settings	FFW	LFW	DW	WSW
All (63)	**35/45**	28/20	16/14	20/17
Grayscale (36)	**26/32**	3/0	3/0	7/4
RGB (18)	6/10	**18/13**	6/7	6/6
DVS (9)	3/3	**7/7**	**7/7**	**7/7**

Figure 6 further shows where these differences occur across datasets, corruption conditions, and memory capacities. FFW is the most frequent winner on grayscale inputs, particularly under noisy cues, whereas LFW becomes more competitive on RGB data. For CIFAR10-DVS, the small winner margins and frequent ties indicate that the choice of temporal readout has less influence on retrieval quality.

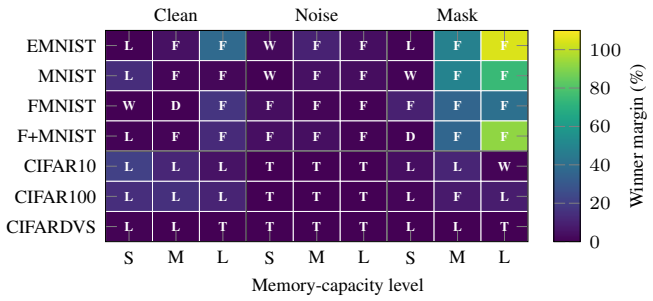


Figure 6: Winner-margin heatmap across datasets, input conditions, and memory capacities. F, L, D, W, and T denote FFW, LFW, DW, WSW, and ties. Colour indicates the winner's MSE improvement over the runner-up; S, M, and L denote the three capacity levels.

Together, these results indicate that the earliest firing winner is often sufficient for retrieval in SHONA, consistent with prior findings that first-to-spike responses can support rapid neural decisions (Gardner & Grüning, 2021; Jin et al., 2002; Thorpe & Gautrais, 1996).

5.5 Robustness and Sensitivity

We further evaluate SHONA's robustness to Gaussian noise and masking, together with its sensitivity to the temporal horizon T, which determines the duration over which its spike-driven retrieval dynamics evolve.

For cue robustness, Table 3a shows that SHONA achieves the lowest mean MSE under masking at both corruption levels, with error increasing by only 8.5% from 0.1053 at 30% masking to 0.1142 at 50%. Under Gaussian noise, SHONA also outperforms MHN by 23.0% at $\sigma = 0.3$ and 6.1% at $\sigma = 0.5$. SQHN and cosine achieve lower MSE under Gaussian noise, but neither shares SHONA's spike-driven Hopfield formulation. Overall, SHONA is particularly robust to incomplete cues.

Within the tested range, Table 3b shows that extending T provides no retrieval benefit. The shortest tested horizon, $T = 300$, achieves the best mean MSE and SSIM $(0.1304, 0.2276)$. Increasing

(a) Cue corruption sensitivity.

	Gaussian Noise		Masking	
Method	$\sigma = 0.3 \downarrow$	$\sigma = 0.5 \downarrow$	$30\% \downarrow$	$50\% \downarrow$
SHONA	0.0950	0.1178	**0.1053**	**0.1142**
MHN	0.1233	0.1254	0.1197	0.1291
SQHN	0.0689	0.0696	0.1247	0.1509
Cosine	0.0658	0.0731	0.1161	0.1364

(b) Temporal horizon sensitivity.

T	Mean MSE $\downarrow$	Mean SSIM $\uparrow$
300	**0.1304**	**0.2276**
350	0.1489	0.2175
400	0.1633	0.1983
$\Delta_{300\to350}$	+14.2%	-4.4%
$\Delta_{300\to400}$	+25.2%	-12.9%

Table 3: Summary of retrieval robustness and temporal sensitivity. (a) Mean MSE under Gaussian noise and masking across 21 dataset–capacity settings; (b) Mean MSE and SSIM across temporal horizons T under Full-CF.

T to 350 degrades MSE by 14.2% and SSIM by 4.4%, while $T = 400$ degrades them by 25.2% and 12.9%, respectively. Thus, additional temporal processing does not improve retrieval over the horizons evaluated. Detailed corruption, SSIM, and dataset-level temporal results are provided in Appendices A.8.1 and A.8.2.

6 Discussion and Limitations

SHONA extends Hopfield-style associative retrieval into an online, spike-driven memory where traces are retrieved and updated. At each winner event, spike-driven retrieval admits a Hopfield-style energy-descent interpretation, and under stated conditions the winner-conditioned update does not increase that event-specific energy. Together with exponential capacity scaling for instantaneous winner selection, this retains Hopfield retrieval principles.

Empirically, the adaptive horizon provides the primary protection against CF by limiting a trace's exposure to plasticity, while dominance pruning provides a complementary safeguard against persistent trace dominance. The temporal experiments show that longer processing is not necessarily beneficial: the shortest tested horizon, $T = 300$, achieves the best overall MSE and SSIM, while FFW performs strongest overall. SHONA is robust under masking, indicating effective retrieval from incomplete cues.

Under Gaussian noise, SQHN and cosine achieve lower MSE, while SHONA remains competitive with the structurally closer MHN, particularly on grayscale datasets. More importantly, SHONA combines Hopfield-style retrieval with spike-mediated dynamics, online trace adaptation, and explicit CF control within a single continual-memory process. These capabilities distinguish its contribution from methods focused primarily on retrieval performance. The experiments are limited to evaluated datasets, memory sizes, corruption levels, and temporal horizons, and energy savings are not measured directly. Future work can extend SHONA to longer, heterogeneous continual streams, deeper or hierarchical memory architectures, broader native event-based tasks, and neuromorphic implementations that evaluate latency, sparsity, and energy efficiency.

7 Conclusion

We introduced SHONA to transition Hopfield AM from a static cue-recovery mechanism into a continual, spike-driven process where new cues update stored traces through timing-dependent plasticity. To mitigate CF from ongoing modification, SHONA regulates spike-mediated trace evolution through an adaptive horizon and dominance pruning, permitting online adaptation. Thus, SHONA unifies retrieval, memory adaptation, and forgetting control within a dynamic spiking AM system.

AI USE STATEMENT

Generative AI tools were used for manuscript editing and presentation, figure preparation, and feedback during experimental analysis. All AI-supported material was independently reviewed and verified by the authors. Experimental data and reported results were obtained from the authors' implementations and experiments. The authors take full responsibility for the content of this work.

REPRODUCIBILITY STATEMENT

The formulation and implementation details of SHONA are provided in the main paper, with additional algorithmic details, experimental settings, dataset and preprocessing information, hyperparameters, ablations, and extended results provided in the appendix. These materials document the settings used for retrieval, online updating, robustness evaluation, and catastrophic-forgetting control to support reproduction of the reported results.

REFERENCES

Nicholas Alonso and Jeffrey L Krichmar. A sparse quantized hopfield network for online-continual memory. *Nature Communications*, 15(1):3722, 2024.

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

Guo-qiang Bi and Mu-ming Poo. Synaptic modifications in cultured hippocampal neurons: Dependence on spike timing, synaptic strength, and postsynaptic cell type. *The Journal of Neuroscience*, 18(24):10464–10472, 1998. doi: 10.1523/JNEUROSCI.18-24-10464.1998.

Guo-Qiang Bi and Jonathan Rubin. Timing in synaptic plasticity: from detection to integration. *Trends in neurosciences*, 28(5):222–228, 2005.

Daniel Casanueva-Morato, Alvaro Ayuso-Martinez, Juan P Dominguez-Morales, Angel Jimenez-Fernandez, and Gabriel Jimenez-Moreno. A bio-inspired implementation of a sparse-learning spike-based hippocampus memory model. *IEEE Transactions on Emerging Topics in Computing*, 13(1):119–133, 2024.

Yang Dan and Mu-Ming Poo. Spike timing-dependent plasticity: from synapse to perception. *Physiological reviews*, 86(3):1033–1048, 2006.

Peter Dayan and Laurence F Abbott. *Theoretical neuroscience: computational and mathematical modeling of neural systems*. MIT press, 2005.

Vincenzo De Paola, Anthony Holtmaat, Graham Knott, Sen Song, Linda Wilbrecht, Pico Caroni, and Karel Svoboda. Cell type-specific structural plasticity of axonal branches and boutons in the adult neocortex. *Neuron*, 49(6):861–875, 2006.

Dominique Debanne, Emilie Campanac, Andrzej Bialowas, Edmond Carlier, and Gisèle Alcaraz. Axon physiology. *Physiological reviews*, 91(2):555–602, 2011.

E Paxon Frady and Friedrich T Sommer. Robust computation with rhythmic spike patterns. *Proceedings of the National Academy of Sciences*, 116(36):18050–18059, 2019.

Brian Gardner and André Grüning. Supervised learning with first-to-spike decoding in multilayer spiking neural networks. *Frontiers in computational neuroscience*, 15:617862, 2021.

Robert Gütig, Ranit Aharonov, Stefan Rotter, and Haim Sompolinsky. Learning input correlations through nonlinear temporally asymmetric hebbian plasticity. *The Journal of neuroscience*, 23(9): 3697–3714, 2003.

Julia Gygax and Friedemann Zenke. Elucidating the theoretical underpinnings of surrogate gradient learning in spiking neural networks. *Neural Computation*, 37(5):886–925, 2025.

Archibald Vivian Hill. The possible effects of the aggregation of the molecules of hemoglobin on its dissociation curves. *j. physiol.*, 40:iv–vii, 1910.

John J Hopfield. Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the national academy of sciences*, 79(8):2554–2558, 1982.

Aidan J. Horner, James A. Bisby, Daniel Bush, Wen-Jing Lin, and Neil Burgess. Evidence for holistic episodic recollection via hippocampal pattern completion. *Nature Communications*, 6: 7462, 2015. doi: 10.1038/ncomms8462.

Dezhe Z Jin, H Sebastian Seung, et al. Fast computation with spikes in a recurrent neural network. *PHYSICAL REVIEW-SERIES E-*, 65(5; PART 1):051922–051922, 2002.

Katsunori Kobayashi and Mu-ming Poo. Spike train timing-dependent associative modification of hippocampal ca3 recurrent synapses by mossy fibers. *Neuron*, 41(3):445–454, 2004.

Dmitry Krotov and John J Hopfield. Dense associative memory for pattern recognition. *Advances in neural information processing systems*, 29, 2016.

Liqun Luo and Dennis DM O'Leary. Axon retraction and degeneration in development and disease. *Annu. Rev. Neurosci.*, 28(1):127–156, 2005.

Henry Markram, Joachim Lübke, Michael Frotscher, and Bert Sakmann. Regulation of synaptic efficacy by coincidence of postsynaptic aps and epsps. *Science*, 275(5297):213–215, 1997. doi: 10.1126/science.275.5297.213.

Colin McDiarmid et al. On the method of bounded differences. *Surveys in combinatorics*, 141(1): 148–188, 1989.

Beren Millidge, Tommaso Salvatori, Yuhang Song, Thomas Lukasiewicz, and Rafal Bogacz. Universal hopfield networks: A general framework for single-shot associative memory models. In *International Conference on Machine Learning*, pp. 15561–15583. PMLR, 2022.

Ryota Miyata, Keisuke Ota, and Toru Aonishi. Optimal design for hetero-associative memory: hippocampal ca1 phase response curve and spike-timing-dependent plasticity. *PLoS One*, 8(10): e77395, 2013.

Abigail Morrison, Markus Diesmann, and Wulfram Gerstner. Phenomenological models of synaptic plasticity based on spike timing. *Biological cybernetics*, 98(6):459–478, 2008.

Su-Hwan Na and Dong-Sun Kim. An energy-efficient neuromorphic processor using unified refractory control-based noc for edge ai. *Electronics*, 14(24):4959, 2025.

Ken-Ichi Naka and William AH Rushton. S-potentials from luminosity units in the retina of fish (cyprinidae). *The Journal of physiology*, 185(3):587–599, 1966.

Hubert Ramsauer, Bernhard Schäfl, Johannes Lehner, Philipp Seidl, Michael Widrich, Thomas Adler, Lukas Gruber, Markus Holzleitner, Milena Pavlović, Geir Kjetil Sandve, et al. Hopfield networks is all you need. *arXiv preprint arXiv:2008.02217*, 2020.

Edmund T. Rolls. The mechanisms for pattern completion and pattern separation in the hippocampus. *Frontiers in Systems Neuroscience*, 7:74, 2013. doi: 10.3389/fnsys.2013.00074.

Amos J Storkey and Romain Valabregue. The basins of attraction of a new hopfield learning rule. *Neural Networks*, 12(6):869–876, 1999.

Simon Thorpe and Jacques Gautrais. Rapid visual processing using spike asynchrony. *Advances in neural information processing systems*, 9, 1996.

Mark CW Van Rossum, Guo Qiang Bi, and Gina G Turrigiano. Stable hebbian learning from spike timing-dependent plasticity. *The Journal of Neuroscience*, 20(23):8812–8821, 2000.

Danying Wang, George Parish, Kimron L Shapiro, and Simon Hanslmayr. Interaction between theta phase and spike timing-dependent plasticity simulates theta-induced memory effects. *Eneuro*, 10 (3), 2023.

Alan L Yuille and Anand Rangarajan. The concave-convex procedure. *Neural computation*, 15(4): 915–936, 2003.

A Appendix

A.1 Periodic Rate Coding

For static image inputs, SHONA represents intensity through spike frequency: larger input values produce denser spike sequences, whereas zero-valued components remain silent. Unlike stochastic Bernoulli or Poisson coding, the encoding is deterministic, so the same input always generates the same spike train.

For an input component p_k, its intensity is normalized within the current image and mapped to the frequency range $[f_{\min}, f_{\max}]$:

$$\bar{p}_k = \frac{p_k - p_{\min}}{p_{\max} - p_{\min}}, \qquad f_k = f_{\min} + \bar{p}_k \left(f_{\max} - f_{\min}\right). \tag{20}$$

The assigned frequency determines the periodic spacing

$$\Delta_k = \left\lceil \frac{T}{f_k} \right\rceil, \qquad x_k(t) = \begin{cases} 1, & p_k > 0 \text{ and } t = m\Delta_k, \quad m \in \mathbb{N}, \\ 0, & \text{otherwise}, \end{cases} \tag{21}$$

for $t = 1, \ldots, T$. Thus, increasing the input intensity shortens the inter-spike interval and provides more frequent evidence to the spike-driven retrieval dynamics. We use $f_{\min} = 1$ and $f_{\max} = 50$ throughout the experiments.

A.2 Spike-based Energy-Guided Retrieval

We show that each winner generated by the spike-driven retrieval mechanism in Equation 12 admits a Hopfield-style energy interpretation. Within an AM iteration ℓ, the input $\boldsymbol{\xi}^{(\ell)}$ is represented by a spike sequence $\boldsymbol{x}^{(\ell)}(t)$ over $t \in \{1, \ldots, T\}$. At spike time t, let $\Xi_{\ell,t} = \{\boldsymbol{\xi}_i\}$ denote the current stored memory bank used for that retrieval event. After a winner is selected, the same winning index is passed to the winner-conditioned online update under the CF mechanism before subsequent retrieval events. The resulting winner events are then combined by the selected temporal readout rule to produce the next AM state $\boldsymbol{\xi}^{(\ell+1)}$.

For each stored trace $\boldsymbol{\xi}_i$, the membrane response at spike time t is

$$u_i^{(\ell)}(t) = u_i^{(\ell)}(t-1) + \boldsymbol{\xi}_i^\top \boldsymbol{x}^{(\ell)}(t) - \eta_u \chi_i^{(\ell)}(t) \tag{22}$$

where the leak indicator is

$$\chi_i^{(\ell)}(t) = \mathbb{I}\left[u_i^{(\ell)}(t-1) + \boldsymbol{\xi}_i^\top \boldsymbol{x}^{(\ell)}(t) > u_{\text{rest}}\right] \tag{23}$$

Define the threshold-crossing margin

$$q_i^{(\ell)}(t) = u_i^{(\ell)}(t) - \vartheta_i^{(\ell)}(t). \tag{24}$$

Conditioning on the membrane history and leak state at the current event, define

$$b_i^{(\ell)}(t) = u_i^{(\ell)}(t-1) - \eta_u \chi_i^{(\ell)}(t), \tag{25}$$

so that

$$u_i^{(\ell)}(t) = b_i^{(\ell)}(t) + \boldsymbol{\xi}_i^\top \boldsymbol{x}^{(\ell)}(t). \tag{26}$$

Let $\mathcal{E}_\ell(t)$ denote the set of traces eligible to compete at time t. The active set is

$$\mathcal{A}_\ell(t) = \left\{ i \in \mathcal{E}_\ell(t) : q_i^{(\ell)}(t) > 0 \right\}, \tag{27}$$

and, when $\mathcal{A}_\ell(t) \neq \varnothing$, the winner is

$$i_\ell^*(t) = \arg\max_{i \in \mathcal{A}_\ell(t)} u_i^{(\ell)}(t) = \arg\max_{i \in \mathcal{A}_\ell(t)} \left[b_i^{(\ell)}(t) + \boldsymbol{\xi}_i^\top \boldsymbol{x}^{(\ell)}(t)\right]. \tag{28}$$

If $\mathcal{A}_\ell(t) = \varnothing$, no winner is selected.

The following theorem formalises the energy interpretation of this instantaneous winner.

Theorem 1 (Spike-based retrieval as a CCCP update)**.** *Fix an associative-memory iteration ℓ and spike time t, and condition on the neuronal state and current memory bank $\boldsymbol{\Xi}_{\ell,t}$ at that event. Suppose $\mathcal{A}_\ell(t) \neq \varnothing$ and the winner $i^*_\ell(t)$ is unique. Then, for the local Hopfield-style energy defined over the current active set, one CCCP step initialized at*

$$\boldsymbol{z}^{(r)} = \boldsymbol{x}^{(\ell)}(t) \tag{29}$$

returns

$$\boldsymbol{z}^{(r+1)} = \boldsymbol{\xi}_{i^*_\ell(t)}, \tag{30}$$

and the corresponding local energy is non-increasing:

$$E_{\ell,t}\left(\boldsymbol{z}^{(r+1)}; \boldsymbol{\Xi}_{\ell,t}\right) \leq E_{\ell,t}\left(\boldsymbol{z}^{(r)}; \boldsymbol{\Xi}_{\ell,t}\right). \tag{31}$$

Proof. For the fixed event (ℓ, t), define

$$F_{\ell,t}\left(\boldsymbol{z}; \boldsymbol{\Xi}_{\ell,t}\right) = \max_{i \in \mathcal{A}_\ell(t)} \left[b_i^{(\ell)}(t) + \boldsymbol{\xi}_i^\top \boldsymbol{z}\right]. \tag{32}$$

The corresponding local Hopfield-style energy is

$$E_{\ell,t}\left(\boldsymbol{z}; \boldsymbol{\Xi}_{\ell,t}\right) = \frac{1}{2}\|\boldsymbol{z}\|^2 - F_{\ell,t}\left(\boldsymbol{z}; \boldsymbol{\Xi}_{\ell,t}\right). \tag{33}$$

Writing

$$E_{\ell,t}(\boldsymbol{z}; \boldsymbol{\Xi}_{\ell,t}) = E_1(\boldsymbol{z}) + E_2(\boldsymbol{z}), \qquad E_1(\boldsymbol{z}) = \frac{1}{2}\|\boldsymbol{z}\|^2, \qquad E_2(\boldsymbol{z}) = -F_{\ell,t}(\boldsymbol{z}; \boldsymbol{\Xi}_{\ell,t}), \tag{34}$$

Because $F_{\ell,t}$ is the maximum of affine functions over the fixed active set $\mathcal{A}_\ell(t)$, it is convex. Hence $E_2(\boldsymbol{z}) = -F_{\ell,t}(\boldsymbol{z}; \boldsymbol{\Xi}_{\ell,t})$ is concave. Accordingly, the CCCP update (Yuille & Rangarajan, 2003) satisfies

$$\nabla E_1\left(\boldsymbol{z}^{(r+1)}\right) \in \partial F_{\ell,t}\left(\boldsymbol{z}^{(r)}; \boldsymbol{\Xi}_{\ell,t}\right). \tag{35}$$

Here, ∂ denotes the subdifferential with respect to $\boldsymbol{z}$.

Set

$$\boldsymbol{z}^{(r)} = \boldsymbol{x}^{(\ell)}(t). \tag{36}$$

Since $i^*_\ell(t)$ is the unique winner,

$$\partial F_{\ell,t}\left(\boldsymbol{z}^{(r)}; \boldsymbol{\Xi}_{\ell,t}\right) = \left\{\boldsymbol{\xi}_{i^*_\ell(t)}\right\}. \tag{37}$$

Because

$$\nabla E_1\left(\boldsymbol{z}^{(r+1)}\right) = \boldsymbol{z}^{(r+1)}, \tag{38}$$

Equation 35 gives

$$\boldsymbol{z}^{(r+1)} = \boldsymbol{\xi}_{i^*_\ell(t)}. \tag{39}$$

By the descent property of CCCP for this convex–concave decomposition,

$$E_{\ell,t}\left(\boldsymbol{z}^{(r+1)}; \boldsymbol{\Xi}_{\ell,t}\right) \leq E_{\ell,t}\left(\boldsymbol{z}^{(r)}; \boldsymbol{\Xi}_{\ell,t}\right). \tag{40}$$

□

For a winner event (ℓ, t), let $\boldsymbol{\Xi}^-_{\ell,t} = \{\boldsymbol{\xi}^-_1, \ldots, \boldsymbol{\xi}^-_M\}$ denote the memory bank used by the retrieval step, and let $\boldsymbol{\Xi}^+_{\ell,t} = \{\boldsymbol{\xi}^+_1, \ldots, \boldsymbol{\xi}^+_M\}$ denote the memory bank after the winner-conditioned update in Equation 14. Only the selected trace $\boldsymbol{\xi}_{i^*_\ell(t)}$ is modified by this plasticity step.

Corollary 2 (Local energy compatibility of the winner-conditioned update)**.** *Suppose the winner-conditioned update is applied at* (ℓ, t) *and* $\sigma \geq 0$*,* $A_- \geq 0$*,* $\rho \leq 1$*, and* $\xi^-_{i^*_\ell(t),k} \leq \xi_{\max}$ *for every* k *with* $x^{(\ell)}_k(t) = 1$*. Then, with the neuronal state and active set at the current winner event held fixed,*

$$E_{\ell,t}\left(\boldsymbol{x}^{(\ell)}(t); \boldsymbol{\Xi}^+_{\ell,t}\right) \leq E_{\ell,t}\left(\boldsymbol{x}^{(\ell)}(t); \boldsymbol{\Xi}^-_{\ell,t}\right). \tag{41}$$

Proof. For every component with $x^{(\ell)}_k(t) = 1$ at the current spike time, the most recent spike in the causal window occurs at the current time, so that

$$\Delta t_k(t) = 0. \tag{42}$$

The corresponding branch of Equation 14 therefore gives

$$\xi^+_{i^*_\ell(t),k} - \xi^-_{i^*_\ell(t),k} = \sigma\left(\xi_{\max} - \xi^-_{i^*_\ell(t),k}\right)^{\mu} A_-(1-\rho) \geq 0. \tag{43}$$

Components for which $x^{(\ell)}_k(t) = 0$ do not contribute to the current inner product. Consequently,

$$\left(\boldsymbol{\xi}^+_{i^*_\ell(t)} - \boldsymbol{\xi}^-_{i^*_\ell(t)}\right)^{\top} \boldsymbol{x}^{(\ell)}(t) \geq 0. \tag{44}$$

Because $i^*_\ell(t)$ is already the maximizing branch at the current winner event, Equation 44 implies

$$F_{\ell,t}\left(\boldsymbol{x}^{(\ell)}(t); \boldsymbol{\Xi}^+_{\ell,t}\right) \geq F_{\ell,t}\left(\boldsymbol{x}^{(\ell)}(t); \boldsymbol{\Xi}^-_{\ell,t}\right). \tag{45}$$

The quadratic term $1/2\|\boldsymbol{x}^{(\ell)}(t)\|^2$ is unchanged, and therefore

$$E_{\ell,t}\left(\boldsymbol{x}^{(\ell)}(t); \boldsymbol{\Xi}^+_{\ell,t}\right) \leq E_{\ell,t}\left(\boldsymbol{x}^{(\ell)}(t); \boldsymbol{\Xi}^-_{\ell,t}\right), \tag{46}$$

which proves Equation 41. □

Equation 41 does not require the online update to be a gradient step on $E_{\ell,t}$. Rather, it shows that an admitted winner-conditioned update cannot increase the current local retrieval energy at the spike pattern that selected that winner. The CF mechanism determines how such online modification proceeds through the adaptive-horizon and dominance-control rules defined in the methodology; no monotonic energy claim is made across a pruning or reset operation.

At any subsequent qualifying winner event, the same local retrieval result applies to the memory bank present at that event. Thus, online update and CF may change the stored memories between events without requiring a single global energy to decrease across the entire online sequence.

Across the winner events in AM iteration ℓ, define

$$\mathcal{W}_\ell = \left(\boldsymbol{\xi}_{i^*_\ell(t)}\right)_{t \in \mathcal{T}_\ell}, \tag{47}$$

where $\mathcal{T}_\ell$ denotes the ordered spike times at which a winner is selected.

The next AM state is obtained through the selected temporal readout,

$$\boldsymbol{\xi}^{(\ell+1)} = \mathcal{R}\left(\mathcal{W}_\ell\right). \tag{48}$$

Here, $\mathcal{R}$ denotes the first-firing winner (FFW), last-firing winner (LFW), most-frequent winner (MFW), or weighted-sum winner (WSW) readout rule where further detail is specified in Appendix A.6. Thus, the temporal readout aggregates winner events generated while the memory state may evolve through winner-conditioned update and CF, and returns the single outer AM state $\boldsymbol{\xi}^{(\ell+1)}$.

Because the membrane history, adaptive thresholds, leak and post-firing states, active memory set, and stored memory state may differ across spike times, $E_{\ell,t}$ is an event-specific local energy rather than a single energy propagated across the spike sequence. Accordingly, Equation 40 establishes energy descent only for the retrieval transition within each fixed qualifying winner event, while

Equation 41 shows that, under its stated conditions, the subsequent winner-conditioned update does not increase that same event-local energy at the spike pattern that selected the winner. These energy statements apply separately at each spike time and neither requires nor implies the stronger condition

$$E_{\ell,t+1} \leq E_{\ell,t} \tag{49}$$

across successive spike times. The CF mechanism regulates online modification and subsequent competition through the adaptive-horizon and dominance-control rules, and no monotonic energy claim is made across a pruning or reset operation.

A.3 High-Probability Exponential Storage Capacity

The following analysis establishes a high-probability exponential-capacity guarantee for the instantaneous winner-selection rule of SHONA The result incorporates membrane relaxation, adaptive firing thresholds, and refractory eligibility while retaining the exponential scaling associated with Hopfield-style associative memories.

Theorem 3 (Exponential capacity of spike-based winner selection)**.** *Fix a spike time t and condition on the preceding neuronal state, including the membrane history, adaptive thresholds, and refractory-eligible memory set. Suppose that the input coordinates are independent, the stored traces are uniformly bounded, and the expected retrieval margin grows linearly with the input dimension d. Then the probability of an incorrect SHONA winner decays exponentially with d. Consequently, the number of stored memories that can be retrieved uniformly with fixed failure probability grows exponentially with d.*

Proof. Let the memory matrix be

$$\boldsymbol{\Xi} = \begin{bmatrix} \boldsymbol{\xi}_1^\top \\ \vdots \\ \boldsymbol{\xi}_N^\top \end{bmatrix} \in \mathbb{R}^{N \times d}, \tag{50}$$

where $\boldsymbol{\xi}_i \in \mathbb{R}^d$ denotes the ith stored trace. Let $\mathcal{H}_{t-1}$ denote the neuronal history preceding spike time t, and let $\mathcal{E}(t)$ contain the memories that are not refractory and may participate in retrieval.

For each eligible memory i, the membrane response is

$$u_i(t) = u_i(t-1) + \boldsymbol{\xi}_i^\top \boldsymbol{x}(t) - \eta_u \chi_i(t), \tag{51}$$

where

$$\chi_i(t) = \mathbb{I}\left[u_i(t-1) + \boldsymbol{\xi}_i^\top \boldsymbol{x}(t) > u_{\text{rest}}\right] \tag{52}$$

denotes whether membrane relaxation is active.

Let $k \in \mathcal{E}(t)$ denote the target memory. A sufficient condition for memory k to be selected is

$$u_k(t) > \max\left\{\vartheta_k(t), \max_{j \in \mathcal{E}(t),\, j \neq k} u_j(t)\right\}, \tag{53}$$

where $\vartheta_k(t)$ is the current adaptive threshold of the target memory. Define the corresponding retrieval margin as

$$\Delta_k(t) = u_k(t) - \max\left\{\vartheta_k(t), \max_{j \in \mathcal{E}(t),\, j \neq k} u_j(t)\right\}. \tag{54}$$

Thus, $\Delta_k(t) > 0$ guarantees that the target is the unique suprathreshold winner, while the failure event is

$$\mathcal{F}_k(t) = \{\Delta_k(t) \leq 0\}. \tag{55}$$

Conditioned on $\mathcal{H}_{t-1}$, the preceding membrane responses, thresholds, and eligible set are fixed. The randomness is therefore carried by the current input vector $\boldsymbol{x}(t) = (x_1(t), \ldots, x_d(t))$ under the assumed input distribution.

Consider two input vectors that differ only in coordinate r. From Equation 51, the resulting change in any eligible membrane response is bounded by

$$|u_i(t; \boldsymbol{x}) - u_i(t; \boldsymbol{x}')| \leq |\xi_{i,r}| + \eta_u, \tag{56}$$

because changing one coordinate alters its synaptic contribution by at most $|\xi_{i,r}|$ and can change the leak indicator by at most one.

Since the maximum operation is non-expansive, the corresponding change in the retrieval margin satisfies

$$|\Delta_k(t;\boldsymbol{x}) - \Delta_k(t;\boldsymbol{x}')| \le c_r, \qquad c_r = 2 \max_{i \in \mathcal{E}(t)} \left(|\xi_{i,r}| + \eta_u\right). \tag{57}$$

Assume that these coordinate-wise bounds satisfy

$$\sum_{r=1}^{d} c_r^2 \le Bd \tag{58}$$

for some constant $B > 0$ independent of d. This condition follows, for example, when the stored-trace coordinates and η_u are uniformly bounded.

Assume further that the conditional expected retrieval margin is bounded below by a positive quantity that grows linearly with dimension:

$$\mathbb{E}\left[\Delta_k(t) \mid \mathcal{H}_{t-1}\right] \ge m_d, \qquad m_d = \zeta d, \tag{59}$$

for some constant $\zeta > 0$ independent of d.

Applying McDiarmid's bounded-difference inequality (McDiarmid et al., 1989) gives

$$\Pr\left(\mathcal{F}_k(t) \mid \mathcal{H}_{t-1}\right) = \Pr\left(\Delta_k(t) \le 0 \mid \mathcal{H}_{t-1}\right) \tag{60}$$

$$\le \exp\left(-\frac{2m_d^2}{\sum_{r=1}^{d} c_r^2}\right) \tag{61}$$

$$\le \exp\left(-\frac{2\zeta^2 d^2}{Bd}\right) = \exp(-\gamma d), \tag{62}$$

where

$$\gamma = \frac{2\zeta^2}{B} > 0. \tag{63}$$

Therefore, for each fixed target memory and neuronal history, the probability of an incorrect winner decays exponentially with the input dimension:

$$\Pr\left(\mathcal{F}_k(t) \mid \mathcal{H}_{t-1}\right) \le \exp(-\gamma d). \tag{64}$$

To obtain a uniform guarantee over the N stored memories, the union bound gives

$$\Pr\left(\bigcup_{k=1}^{N} \mathcal{F}_k(t) \,\middle|\, \mathcal{H}_{t-1}\right) \le N \exp(-\gamma d). \tag{65}$$

Hence, the probability that at least one target fails to satisfy the winner-selection condition is at most δ whenever

$$N \exp(-\gamma d) \le \delta, \tag{66}$$

or equivalently,

$$N \le \delta \exp(\gamma d). \tag{67}$$

Thus, provided that the bounded-difference and linear-margin assumptions hold uniformly for all targets and all memory sizes up to

$$C_\delta(d) = \lfloor \delta \exp(\gamma d) \rfloor, \tag{68}$$

SHONA admits a certified instantaneous winner-selection capacity satisfying

$$C_\delta(d) = \Omega\left(\exp(\gamma d)\right). \tag{69}$$

□

The theorem applies to the winner selected at a fixed spike time under the current neuronal state. The first-winner, last-winner, dominant-winner, and weighted readouts subsequently aggregate winner events over the spike horizon and are not required for this instantaneous capacity result.

A.4 Adaptive Horizon Signal Definitions

The adaptive horizon is driven by three signals that describe how the winning traces evolve during online operation. First, the interval between successive winner events is

$$\iota_k = t_k - t_{k-1}, \tag{70}$$

where t_k and t_{k-1} denote consecutive winner-event times. This interval provides a temporal measure of update activity, distinguishing closely spaced winner events from more widely separated ones.

The selection count of trace i is

$$c_{i,k} = \sum_{r=1}^{k} \mathbb{I}[i_r = i], \tag{71}$$

where i_r denotes the winning trace at event r. Thus, $c_{i,k}$ records how often trace i has been selected and captures repeated exposure of the same memory to winner-conditioned updates.

To retain this information over a longer history, SHONA maintains the accumulated usage

$$a_i \leftarrow a_i + \frac{c_i}{(\max_j c_j + \epsilon)}, \tag{72}$$

where the winner count is normalised against the most frequently selected trace and $\epsilon > 0$ prevents division by zero. Unlike the instantaneous event spacing and current selection count, a_i accumulates sustained reliance on trace i over time.

Taken together, ι_k, $c_{i,k}$, and a_i capture the temporal concentration of winner events, repeated selection of a particular trace, and its longer-term usage, respectively. These complementary signals are used in Equation 17 to regulate how much of the original update horizon remains available for further plasticity.

A.5 Misassignment-Induced Trace Corruption and Bounded Exposure Under Dominance Pruning

Dominance pruning addresses a failure mode that arises from combining winner-based retrieval with online plasticity. The SHONA winner rule measures operational similarity through membrane responses. Consequently, a cue may produce the strongest suprathreshold response for a memory even when it is not semantically associated with the content represented by that memory. Because plasticity is winner-conditioned, such a false-positive assignment modifies the selected trace.

We formalise this process by distinguishing the operational response basin of a memory from the set of cues that are semantically valid for that memory. We then show that repeated false-positive assignments can produce measurable trace corruption, while dominance pruning places a finite bound on the number of such dominant assignments that can occur before reset.

A.5.1 Operational and semantic assignment

Let ℓ index input episodes in the online stream, and let $\boldsymbol{\xi}_{\text{in}}^{(\ell)}$ be encoded as the spike train $\boldsymbol{x}^{(\ell)}(t)$ over the effective horizon $t \in \{1, \ldots, T_\ell\}$. Let $\mathcal{E}^{(\ell)}(t)$ denote the memories that are eligible to participate at spike time t after accounting for refractory state.

The instantaneous response basin of memory p is

$$\boldsymbol{x}^{(\ell)}(t) \in \mathcal{B}_p^{(\ell)}(t) \iff \begin{cases} p \in \mathcal{E}^{(\ell)}(t), \\ u_p^{(\ell)}(t) > \vartheta_p^{(\ell)}(t), \\ u_p^{(\ell)}(t) > u_j^{(\ell)}(t), \quad j \in \mathcal{E}^{(\ell)}(t) \setminus \{p\}, \end{cases} \tag{73}$$

where $u_i^{(\ell)}(t)$ follows the membrane-response dynamics in Equation 11a, and $\vartheta_i^{(\ell)}(t)$ follows the adaptive-threshold dynamics in Equation 11b. Therefore, by Equation 12,

$$\boldsymbol{x}^{(\ell)}(t) \in \mathcal{B}_p^{(\ell)}(t) \quad \Longrightarrow \quad i_\ell^*(t) = p \tag{74}$$

Let the winner count of memory i during episode ℓ be

$$c_i^{(\ell)} = \sum_{t=1}^{T_\ell} \mathbb{I}\left[i_\ell^*(t) = i\right], \tag{75}$$

where spike times at which no winner is selected contribute zero.

The repeated-selection statistic of memory i is updated across input episodes as

$$s_i^{(\ell)} = \begin{cases} s_i^{(\ell-1)} + \dfrac{c_i^{(\ell)}}{\left(\max_j c_j^{(\ell)} + \epsilon\right)}, & c_i^{(\ell)} > 0, \\ 0, & c_i^{(\ell)} = 0. \end{cases} \tag{76}$$

Thus, $s_i^{(\ell)}$ accumulates while memory i continues to reappear among the winners from one input episode to the next, and resets when it receives no winner event in an episode.

The dominant memory for the episode is

$$i_\ell^\dagger = \arg\max_i c_i^{(\ell)}. \tag{77}$$

For each memory p, let $\mathcal{C}_p$ denote the set of cues that are semantically valid for the content represented by $\boldsymbol{\xi}_p$. The set $\mathcal{C}_p$ may be defined by pattern identity, class membership, or another task-level validity relation. It is used only for analysis and need not be available to SHONA during online operation.

The false-positive dominant-assignment region is

$$\mathcal{O}_p^{(\ell)} = \left\{ \boldsymbol{\xi}_{\text{in}}^{(\ell)} : i_\ell^\dagger = p,\ c_p^{(\ell)} > 0,\ \boldsymbol{\xi}_{\text{in}}^{(\ell)} \notin \mathcal{C}_p \right\}. \tag{78}$$

Therefore, a cue in $\mathcal{O}_p^{(\ell)}$ is operationally assigned to memory p even though it is not semantically valid for that memory.

A.5.2 Misassignment-induced trace corruption

Lemma 4 (Off-target assignment and trace corruption)**.** *Suppose memory p receives a sequence of K winner-conditioned updates generated by cues outside $\mathcal{C}_p$. Let $\boldsymbol{\delta}_p^{(r)}$ denote the vector update generated by the rth such winner event under Equation 14 with the spike timing defined by Equations 15a–15b. Then the displacement of the trace from its state*

$$\boldsymbol{e}_p(K) = \boldsymbol{\xi}_p^{(K)} - \boldsymbol{\xi}_p^{(0)} = \sum_{r=0}^{K-1} \boldsymbol{\delta}_p^{(r)}. \tag{79}$$

Define the representation-corruption measure

$$\mathcal{Q}_p(K) = \left\|\boldsymbol{e}_p(K)\right\|^2. \tag{80}$$

Let $\mathscr{F}_r$ denote the update history preceding the rth off-target update. Suppose there exists $\nu > 0$ such that

$$\mathbb{E}\left[\left\langle \boldsymbol{e}_p(r), \boldsymbol{\delta}_p^{(r)} \right\rangle \middle| \mathscr{F}_r\right] \geq 0, \tag{81}$$

$$\mathbb{E}\left[\left\|\boldsymbol{\delta}_p^{(r)}\right\|^2 \middle| \mathscr{F}_r\right] \geq \nu. \tag{82}$$

Then

$$\mathbb{E}\left[\mathcal{Q}_p(K)\right] \geq K\nu. \tag{83}$$

Hence, under non-cancelling, non-trivial off-target plasticity, the expected squared displacement of the stored trace grows at least linearly with the number of false-positive updates.

Proof. Because the SHONA update is winner-conditioned, every event for which $i_\ell^*(t) = p$ modifies only the selected trace. In vector form, the rth update is

$$\boldsymbol{\xi}_p^{(r+1)} = \boldsymbol{\xi}_p^{(r)} + \boldsymbol{\delta}_p^{(r)}. \tag{84}$$

Expanding this recursion gives Equation 79.

The corruption measure after the next update satisfies

$$\mathcal{Q}_p(r+1) = \left\| \boldsymbol{e}_p(r) + \boldsymbol{\delta}_p^{(r)} \right\|^2 \tag{85}$$

$$= \mathcal{Q}_p(r) + 2\left\langle \boldsymbol{e}_p(r), \boldsymbol{\delta}_p^{(r)} \right\rangle + \left\| \boldsymbol{\delta}_p^{(r)} \right\|^2. \tag{86}$$

Taking the conditional expectation and applying Equations 81 and 82 yields

$$\mathbb{E}\left[\mathcal{Q}_p(r+1) | \mathscr{F}_r\right] \geq \mathcal{Q}_p(r) + \nu. \tag{87}$$

Iterating from $\mathcal{Q}_p(0) = 0$ gives

$$\mathbb{E}\left[\mathcal{Q}_p(K)\right] \geq K\nu. \tag{88}$$

□

The non-cancellation condition does not require all off-target updates to point in the same direction. It only excludes the special case in which successive updates systematically reverse the existing displacement. For example, the condition holds for conditionally uncorrelated, zero-mean update increments with non-zero conditional variance.

A.5.3 Bounded corruption exposure under dominance pruning

Theorem 5 (Bounded off-target exposure under dominance pruning)**.** *For episode ℓ, define the normalised exposure of memory i as*

$$\alpha_i^{(\ell)} = \frac{c_i^{(\ell)}}{\max_j c_j^{(\ell)} + \varepsilon}. \tag{89}$$

Suppose that memory p is dominant whenever it is counted in the following bounds, and assume

$$a_p(0) \leq \tau_a, \qquad s_p(0) \leq \tau_s. \tag{90}$$

Under the SHONA reset condition

$$a_{i^\dagger} > \tau_a \quad \textit{or} \quad s_{i^\dagger} > \tau_s, \tag{91}$$

the number of consecutive dominant episodes received by p before reset is bounded by

$$L_{\max} = \lfloor (1+\varepsilon)\left(\tau_s - s_p(0)\right) \rfloor + 1, \tag{92}$$

and the total number of dominant pruning checks, whether consecutive or not, is bounded by

$$R_{\max} = \lfloor (1+\varepsilon)\left(\tau_a - a_p(0)\right) \rfloor + 1. \tag{93}$$

If each input episode contains at most T winner events and each plasticity increment satisfies

$$\left\| \boldsymbol{\delta}_p^{(r)} \right\| \leq M \tag{94}$$

for some finite M, then the off-target corruption accumulated during repeated dominant episodes before reset satisfies

$$\mathcal{Q}_p \leq M^2 T^2 R_{\max}^2. \tag{95}$$

For a consecutive dominant run, the sharper bound is

$$\mathcal{Q}_p \leq M^2 T^2 \min\left\{L_{\max}, R_{\max}\right\}^2. \tag{96}$$

Therefore, dominance pruning prevents a false-positive dominant memory from receiving an unbounded number of off-target updates.

Proof. If $p = i_\ell^\dagger$ and $c_p^{(\ell)} \geq 1$, then

$$\alpha_p^{(\ell)} = \frac{c_p^{(\ell)}}{\max_j c_j^{(\ell)} + \varepsilon} \tag{97}$$

$$= \frac{c_p^{(\ell)}}{c_p^{(\ell)} + \varepsilon} \tag{98}$$

$$\geq \frac{1}{1+\varepsilon}. \tag{99}$$

Suppose first that p remains dominant for L consecutive input episodes. The consecutive-use statistic therefore satisfies

$$s_p(L) \geq s_p(0) + \frac{L}{1+\varepsilon}. \tag{100}$$

Taking $L = L_{\max}$ from Equation 92 gives

$$\frac{L_{\max}}{1+\varepsilon} > \tau_s - s_p(0), \tag{101}$$

and hence

$$s_p(L_{\max}) > \tau_s. \tag{102}$$

The streak component of the reset condition is therefore reached after at most $L_{\max}$ consecutive dominant episodes.

For dominant episodes that are not necessarily consecutive, the accumulated usage update in Equation 16 and Equation 99 give

$$a_p(R) \geq a_p(0) + \frac{R}{1+\varepsilon} \tag{103}$$

after R dominant pruning checks. Taking $R = R_{\max}$ from Equation 93 yields

$$a_p(R_{\max}) > \tau_a. \tag{104}$$

Thus, the accumulated-use component triggers reset after at most $R_{\max}$ dominant pruning checks.

Because at most one winner is selected at each internal spike time, memory p can receive no more than T winner-conditioned updates during one input episode. Therefore, before the accumulated-use reset,

$$K \leq T R_{\max}. \tag{105}$$

Using Equation 79 and the triangle inequality,

$$\mathcal{Q}_p(K) = \left\| \sum_{r=0}^{K-1} \boldsymbol{\delta}_p^{(r)} \right\|^2 \tag{106}$$

$$\leq \left(\sum_{r=0}^{K-1} \left\| \boldsymbol{\delta}_p^{(r)} \right\| \right)^2 \tag{107}$$

$$\leq K^2 M^2 \tag{108}$$

$$\leq M^2 T^2 R_{\max}^2. \tag{109}$$

During a consecutive dominant run, reset occurs when either the streak or accumulated-use threshold is first crossed, giving

$$K \leq T \min\left\{ L_{\max}, R_{\max} \right\}, \tag{110}$$

which proves Equation 96.

□

Lemma 4 and Theorem 5 establish complementary parts of the pruning argument. A cue can be operationally similar enough to dominate the membrane-response competition while remaining semantically invalid for the selected trace. Such a false-positive assignment applies an off-target plasticity update, and repeated non-cancelling updates increase the representation displacement of that trace. Dominance pruning does not detect semantic invalidity directly and does not reconstruct the original memory. Rather, it prevents a repeatedly dominant, potentially corrupted trace from absorbing off-target updates indefinitely and releases the slot through reset.

A.6 Temporal Spike-Aggregated Readout Strategies

The spike-driven retrieval dynamics may activate different stored memories at different spike times. We therefore present four readout strategies that convert the resulting sequence of winner-firing events into a final retrieved representation: first-firing winner (FFW), last-firing winner (LFW), most-frequent winner (MFW), and weighted-sum winner (WSW). These strategies operate on the same membrane-response dynamics and differ only in how the winner events are aggregated over the retrieval horizon. The individual winner events are intermediate retrieval outcomes; the selected readout determines the final retrieved representation.

Let the stored memory matrix be

$$\boldsymbol{\Xi} = \begin{bmatrix} \boldsymbol{\xi}_1^\top \\ \vdots \\ \boldsymbol{\xi}_N^\top \end{bmatrix} \in \mathbb{R}^{N \times d}, \tag{111}$$

where $\boldsymbol{\xi}_i \in \mathbb{R}^d$ denotes the complete memory stored in slot i, and $\xi_{i,k}$ denotes its kth component.

At each spike time t, the membrane responses and adaptive thresholds evolve according to Equations 11a and 11b. Let $\mathcal{E}(t)$ denote the set of memories eligible to participate in retrieval after accounting for their refractory states. The set of eligible memories whose responses exceed their current thresholds is

$$\mathcal{A}(t) = \{i \in \mathcal{E}(t) : u_i(t) > \vartheta_i(t)\}. \tag{112}$$

Whenever $\mathcal{A}(t) \neq \varnothing$, the instantaneous winner is

$$i^*(t) = \arg\max_{i \in \mathcal{A}(t)} u_i(t), \tag{113}$$

as defined by the SHONA winner rule in Equation 12. When $\mathcal{A}(t) = \varnothing$, no winner-firing event is recorded at that spike time.

Suppose that the retrieval process produces K winner-firing events at times

$$t_1 < t_2 < \cdots < t_K, \tag{114}$$

with the corresponding winner sequence

$$\mathcal{W} = (i_1, i_2, \ldots, i_K), \qquad i_r = i^*(t_r). \tag{115}$$

The total number of times memory i wins during the retrieval horizon is

$$c_i = \sum_{r=1}^{K} \mathbb{I}[i_r = i], \tag{116}$$

where $\mathbb{I}[\cdot]$ is the indicator function. The set of memories that produce at least one winner-firing event is

$$\mathcal{V} = \{i : c_i > 0\}. \tag{117}$$

The notation $\mathcal{V}$ distinguishes the memories that participate over the complete retrieval horizon from the instantaneous suprathreshold set $\mathcal{A}(t)$.

A.6.1 First-Firing Winner

The FFW readout selects the memory associated with the earliest winner-firing event:

$$i_{\mathrm{FFW}} = i_1, \tag{118}$$

and the retrieved representation is

$$\widehat{\boldsymbol{\xi}}_{\mathrm{FFW}} = \boldsymbol{\xi}_{i_{\mathrm{FFW}}}. \tag{119}$$

FFW therefore uses the earliest decisive threshold-crossing response. Once the first event occurs at time t_1, its retrieved memory is determined and is not changed by subsequent winner-firing events.

A.6.2 Last-Firing Winner

The LFW readout selects the memory associated with the final winner-firing event:

$$i_{\mathrm{LFW}} = i_K, \tag{120}$$

with retrieval defined as

$$\widehat{\boldsymbol{\xi}}_{\mathrm{LFW}} = \boldsymbol{\xi}_{i_{\mathrm{LFW}}}. \tag{121}$$

Unlike FFW, LFW reflects the memory selected after the full temporal competition has unfolded. It therefore captures the final winner produced by the spike-driven retrieval trajectory.

A.6.3 Most Frequent Winner

The MFW readout selects the memory with the largest total number of winner-firing events:

$$i_{\mathrm{MFW}} = \arg\max_{i \in \mathcal{V}} c_i. \tag{122}$$

The retrieved representation is then

$$\widehat{\boldsymbol{\xi}}_{\mathrm{MFW}} = \boldsymbol{\xi}_{i_{\mathrm{MFW}}}. \tag{123}$$

MFW discards the temporal order of the events and instead uses winner frequency as evidence of sustained cue–memory similarity. A memory that repeatedly wins the thresholded competition is therefore preferred over memories that win only occasionally.

A.6.4 Weighted-Sum Winner

The WSW readout combines all memories that participate in the winner-firing trajectory. Their winner counts are first normalised using a softmax:

$$\alpha_i = \frac{\exp(c_i)}{\sum_{j \in \mathcal{V}} \exp(c_j)}, \qquad i \in \mathcal{V}. \tag{124}$$

The retrieved representation is then

$$\widehat{\boldsymbol{\xi}}_{\mathrm{WSW}} = \sum_{i \in \mathcal{V}} \alpha_i \boldsymbol{\xi}_i. \tag{125}$$

WSW differs from the three hard-selection strategies because it forms a composite representation in which memories with larger winner counts receive greater weight.

The four strategies capture complementary forms of temporal evidence. FFW and LFW depend on the ordering of winner events, whereas DW and WSW depend on their accumulated frequencies. FFW, LFW, and DW are hard readouts that return one stored trace, while WSW is a soft readout that interpolates among the memories activated during retrieval.

If no memory crosses its adaptive threshold during the retrieval horizon, then $K = 0$. In this case, the implementation falls back to the memory with the largest final membrane response:

$$i_0 = \arg \max_{i \in \{1,\dots,N\}} u_i(T). \tag{126}$$

All hard-selection strategies return $\boldsymbol{\xi}_{i_0}$. For WSW, the participating set contains only i_0 with unit weight. Consequently,

$$\widehat{\boldsymbol{\xi}}_{\mathrm{FFW}} = \widehat{\boldsymbol{\xi}}_{\mathrm{LFW}} = \widehat{\boldsymbol{\xi}}_{\mathrm{DW}} = \widehat{\boldsymbol{\xi}}_{\mathrm{WSW}} = \boldsymbol{\xi}_{i_0}. \tag{127}$$

A.7 Extended Experimental Results

A.7.1 Extended Associative Retrieval Results

To complement the MSE results in Section 5.2, we use SSIM to measure structural preservation during retrieval. The evaluation follows the same datasets, memory sizes, cue conditions, and FFW readout used in the main retrieval experiments, with results shown in Table 4.

Table 4: Online retrieval SSIM using FFW readout. Each entry reports *Clean / Noise / Mask*. Noise uses Gaussian noise with $\sigma = 0.5$, and Mask removes 50% of the cue.

Dataset	Size	SHONA	MHN	SQHN	Cosine
MNIST	300	0.3248 / 0.3137 / 0.2880	0.2754 / 0.2299 / 0.2996	0.3492 / 0.3346 / 0.2409	0.3349 / 0.2688 / 0.3130
	1300	0.5977 / 0.2854 / 0.5313	0.2575 / 0.2434 / 0.2542	0.6659 / 0.6362 / 0.3407	0.6555 / 0.5003 / 0.5421
	2300	0.6309 / 0.2684 / 0.5264	0.2634 / 0.2468 / 0.2462	0.7041 / 0.6675 / 0.2631	0.6975 / 0.5528 / 0.5354
EMNIST	300	0.3241 / 0.2638 / 0.2707	0.2226 / 0.1566 / 0.1885	0.3596 / 0.3489 / 0.2376	0.2812 / 0.1580 / 0.2350
	1300	0.5565 / 0.2962 / 0.4478	0.2342 / 0.1498 / 0.2006	0.6322 / 0.6146 / 0.3375	0.6173 / 0.5004 / 0.3710
	2300	0.5668 / 0.2235 / 0.4477	0.1659 / 0.1146 / 0.1586	0.6725 / 0.6483 / 0.3215	0.6628 / 0.5474 / 0.3679
FMNIST	300	0.2962 / 0.3526 / 0.2780	0.3443 / 0.3105 / 0.2042	0.4167 / 0.4148 / 0.2287	0.4154 / 0.3620 / 0.2238
	1300	0.3883 / 0.2120 / 0.3239	0.1906 / 0.1371 / 0.1772	0.6044 / 0.5892 / 0.3116	0.6005 / 0.4912 / 0.2828
	2300	0.4235 / 0.2015 / 0.3248	0.2014 / 0.1786 / 0.1960	0.6605 / 0.6365 / 0.3115	0.6532 / 0.5380 / 0.3125
F+MNIST	300	0.2925 / 0.2230 / 0.2745	0.2195 / 0.1963 / 0.1768	0.3603 / 0.3420 / 0.1863	0.3509 / 0.2856 / 0.1670
	1300	0.4287 / 0.1848 / 0.3363	0.1551 / 0.1328 / 0.1072	0.5821 / 0.5583 / 0.3565	0.5805 / 0.4580 / 0.2084
	2300	0.3684 / 0.1305 / 0.3462	0.1336 / 0.1253 / 0.1305	0.6445 / 0.6111 / 0.3278	0.6296 / 0.5137 / 0.2248
CIFAR10	700	0.1343 / 0.1378 / 0.1259	0.1503 / 0.1511 / 0.1042	0.1811 / 0.1736 / 0.0853	0.1859 / 0.1749 / 0.0805
	2000	0.1942 / 0.3225 / 0.1839	0.3555 / 0.3564 / 0.2085	0.3578 / 0.3258 / 0.1177	0.3574 / 0.3424 / 0.1355
	3300	0.2456 / 0.3843 / 0.2387	0.4062 / 0.4087 / 0.2829	0.4005 / 0.3642 / 0.1339	0.3997 / 0.3825 / 0.1287
CIFAR100	700	0.1320 / 0.1275 / 0.1212	0.1302 / 0.1294 / 0.0994	0.1521 / 0.1458 / 0.0862	0.1628 / 0.1449 / 0.0976
	2000	0.2019 / 0.3242 / 0.1882	0.3508 / 0.3536 / 0.2375	0.3616 / 0.3369 / 0.0987	0.3605 / 0.3513 / 0.1171
	3300	0.2529 / 0.3906 / 0.2244	0.4232 / 0.4257 / 0.2208	0.4087 / 0.3719 / 0.0875	0.4087 / 0.3891 / 0.0861
CIFARDVS	150	0.0928 / 0.1055 / 0.0879	0.1066 / 0.1029 / 0.0933	0.0919 / 0.0917 / 0.0769	0.1120 / 0.1091 / 0.0858
	300	0.0994 / 0.1267 / 0.1019	0.1245 / 0.1199 / 0.1162	0.1390 / 0.1393 / 0.1204	0.1443 / 0.1410 / 0.1088
	500	0.0979 / 0.1171 / 0.1010	0.1115 / 0.1046 / 0.1070	0.1470 / 0.1470 / 0.0995	0.1421 / 0.1396 / 0.1067

The additional evaluation against KNN and Manhattan L_1 provides a deeper analysis of nearest-neighbour, non-Hopfield baselines that omit the recurrent retrieval dynamics used by SHONA. As shown in Table 5, SHONA is particularly robust under masking, outperforming KNN in $15/21$ and L_1 in $19/21$ masked evaluations.

In these evaluations, SHONA achieves a mean masked MSE of 0.1142, compared with 0.1400 for KNN and 0.1470 for L_1. Thus, although direct nearest-neighbour matching remains strong when most cue information is available, SHONA is less sensitive to substantial cue removal.

A.7.2 Extended Retention and Forgetting Control Results

Table 6 provides the complete retention results underlying the summary in Section 5.3. Each setting evaluates the same early set of 50 memories after continued online updates under Full-CF, No-CF, Horizon-only, and Pruning-only. Results are shown across all datasets, memory sizes, and clean, noisy, and masked cues.

Full-CF achieves lower retention MSE than No-CF in $50/63$ settings. The 13 exceptions are concentrated under noisy cues, which account for $11/13$, compared with only one clean and one masked setting. The benefit is particularly clear on CIFAR10-DVS: at size 300, Full-CF reduces MSE from

Table 5: Retrieval comparison with nearest-neighbour baselines under Full-CF. Each entry reports *Clean / Noise / Mask* MSE. Noise uses Gaussian noise with $\sigma = 0.5$, and Mask removes 50% of the cue. Lower MSE is better.

Dataset	Size	SHONA	KNN	L_1
MNIST	300	0.1389 / 0.1090 / 0.1459	0.0951 / 0.0997 / 0.1142	0.1019 / 0.0972 / 0.1559
	1300	0.0570 / 0.1043 / 0.0631	0.0298 / 0.0380 / 0.0753	0.0347 / 0.0330 / 0.1010
	2300	0.0480 / 0.1239 / 0.0625	0.0265 / 0.0345 / 0.0813	0.0309 / 0.0297 / 0.0960
EMNIST	300	0.1414 / 0.1237 / 0.1509	0.1084 / 0.1138 / 0.1450	0.1132 / 0.1111 / 0.1784
	1300	0.0618 / 0.1037 / 0.0812	0.0419 / 0.0495 / 0.1112	0.0483 / 0.0451 / 0.1179
	2300	0.0663 / 0.1780 / 0.0859	0.0372 / 0.0464 / 0.1080	0.0423 / 0.0405 / 0.1088
FMNIST	300	0.1453 / 0.1125 / 0.1681	0.0890 / 0.0956 / 0.1855	0.0986 / 0.0906 / 0.1994
	1300	0.0882 / 0.1644 / 0.1120	0.0320 / 0.0433 / 0.1404	0.0372 / 0.0343 / 0.1373
	2300	0.0937 / 0.2220 / 0.1203	0.0293 / 0.0452 / 0.1374	0.0339 / 0.0323 / 0.1308
F+MNIST	300	0.1571 / 0.1381 / 0.1619	0.1068 / 0.1130 / 0.1552	0.1181 / 0.1089 / 0.1758
	1300	0.0773 / 0.1446 / 0.1168	0.0367 / 0.0452 / 0.1048	0.0426 / 0.0400 / 0.1126
	2300	0.1143 / 0.2612 / 0.1177	0.0314 / 0.0421 / 0.1081	0.0360 / 0.0347 / 0.1021
CIFAR10	700	0.1237 / 0.0540 / 0.1292	0.0516 / 0.0556 / 0.1510	0.0524 / 0.0523 / 0.1361
	2000	0.0385 / 0.0174 / 0.0405	0.0130 / 0.0374 / 0.1513	0.0136 / 0.0138 / 0.1527
	3300	0.0282 / 0.0186 / 0.0267	0.0116 / 0.0342 / 0.0916	0.0127 / 0.0128 / 0.0982
CIFAR100	700	0.1278 / 0.0589 / 0.1342	0.0564 / 0.0581 / 0.1859	0.0573 / 0.0573 / 0.1624
	2000	0.0387 / 0.0186 / 0.0420	0.0133 / 0.0344 / 0.1339	0.0140 / 0.0143 / 0.1097
	3300	0.0308 / 0.0203 / 0.0324	0.0118 / 0.0316 / 0.1256	0.0129 / 0.0131 / 0.1252
CIFAR10-DVS	150	0.2722 / 0.2523 / 0.2814	0.2513 / 0.2522 / 0.2783	0.2522 / 0.2514 / 0.2993
	300	0.1613 / 0.1181 / 0.1627	0.1123 / 0.1131 / 0.1857	0.1199 / 0.1125 / 0.2029
	500	0.1667 / 0.1312 / 0.1618	0.0994 / 0.1020 / 0.1698	0.1108 / 0.0997 / 0.1837

Table 6: Retention MSE of early stored memories under continual updates using FFW readout. Each entry reports *Clean / Noise / Mask*; lower is better. Size denotes the number of stored memories. Noise uses $\sigma = 0.5$, and Mask removes 50% of the cue.

Dataset	Size	Full-CF	No-CF	Horizon	Pruning
MNIST	300	0.1146 / 0.1008 / 0.1210	0.1361 / 0.1305 / 0.1460	0.1147 / 0.1001 / 0.1200	0.1359 / 0.1269 / 0.1450
	1300	0.0198 / 0.1259 / 0.0371	0.0577 / 0.0998 / 0.0768	0.0198 / 0.1259 / 0.0371	0.0577 / 0.0998 / 0.0768
	2300	0.0089 / 0.1175 / 0.0142	0.0173 / 0.1050 / 0.0205	0.0089 / 0.1175 / 0.0142	0.0173 / 0.1050 / 0.0205
EMNIST	300	0.1186 / 0.1092 / 0.1366	0.1561 / 0.1503 / 0.1690	0.1209 / 0.1064 / 0.1375	0.1537 / 0.1457 / 0.1671
	1300	0.0228 / 0.0906 / 0.0407	0.0682 / 0.0570 / 0.0834	0.0228 / 0.0906 / 0.0407	0.0682 / 0.0570 / 0.0834
	2300	0.0193 / 0.1680 / 0.0503	0.0233 / 0.1072 / 0.0426	0.0193 / 0.1680 / 0.0503	0.0233 / 0.1072 / 0.0426
FMNIST	300	0.1256 / 0.0983 / 0.1731	0.1969 / 0.1632 / 0.2235	0.1301 / 0.0950 / 0.1664	0.1921 / 0.1573 / 0.2245
	1300	0.0528 / 0.1669 / 0.0985	0.0812 / 0.0744 / 0.1554	0.0528 / 0.1669 / 0.0985	0.0812 / 0.0744 / 0.1554
	2300	0.0480 / 0.2362 / 0.0985	0.0559 / 0.1478 / 0.1342	0.0480 / 0.2362 / 0.0985	0.0559 / 0.1478 / 0.1342
F+MNIST	300	0.1384 / 0.1146 / 0.1425	0.1744 / 0.1466 / 0.1815	0.1435 / 0.1184 / 0.1446	0.1778 / 0.1489 / 0.1840
	1300	0.0441 / 0.1629 / 0.0822	0.0867 / 0.0787 / 0.1214	0.0441 / 0.1629 / 0.0822	0.0867 / 0.0787 / 0.1214
	2300	0.0379 / 0.2736 / 0.0758	0.0399 / 0.2058 / 0.0849	0.0379 / 0.2736 / 0.0758	0.0399 / 0.2058 / 0.0849
CIFAR10	700	0.1170 / 0.0509 / 0.1143	0.1092 / 0.0543 / 0.1342	0.1184 / 0.0519 / 0.1146	0.1007 / 0.0443 / 0.1146
	2000	0.0364 / 0.0221 / 0.0356	0.0634 / 0.0293 / 0.0556	0.0364 / 0.0221 / 0.0356	0.0567 / 0.0273 / 0.0629
	3300	0.0176 / 0.0245 / 0.0216	0.0266 / 0.0224 / 0.0300	0.0176 / 0.0245 / 0.0216	0.0266 / 0.0224 / 0.0300
CIFAR100	700	0.1217 / 0.0547 / 0.1249	0.1334 / 0.0623 / 0.1324	0.1234 / 0.0610 / 0.1257	0.1289 / 0.0618 / 0.1405
	2000	0.0365 / 0.0253 / 0.0377	0.0628 / 0.0320 / 0.0673	0.0365 / 0.0253 / 0.0377	0.0628 / 0.0320 / 0.0673
	3300	0.0195 / 0.0287 / 0.0190	0.0402 / 0.0264 / 0.0383	0.0195 / 0.0287 / 0.0190	0.0402 / 0.0264 / 0.0383
CIFARDVS	150	0.1774 / 0.1482 / 0.1764	0.3053 / 0.2918 / 0.3145	0.1774 / 0.1482 / 0.1764	0.3053 / 0.2918 / 0.3145
	300	0.0872 / 0.0444 / 0.0719	0.2379 / 0.1940 / 0.2517	0.0872 / 0.0444 / 0.0719	0.2379 / 0.1940 / 0.2517
	500	0.0759 / 0.1556 / 0.0674	0.1870 / 0.1087 / 0.1877	0.0759 / 0.1556 / 0.0674	0.1870 / 0.1087 / 0.1877

0.2379 to 0.0872 for clean cues, from 0.1940 to 0.0444 under noise, and from 0.2517 to 0.0719 under masking.

Table 7a further aggregates these results to separate the effects of the two CF components. Horizon-only matches Full-CF in $45/63$ retention settings and remains similarly close for online MSE and SSIM, while Table 7b shows Full-CF outperforming Pruning-only in $49/63$ retention settings, alongside $47/63$ for both online MSE and SSIM in Table 7a. This supports the complementary roles observed in the main results: the horizon provides primary protection against progressive overwrite, while pruning provides corrective control over persistent trace dominance.

Table 7: Summary of CF effects across 63 dataset, memory-size, and cue settings. (a) Full-CF relative to each ablation for retention MSE, online MSE, and SSIM, reported as *better / equal / worse*. (b) Retention MSE comparison against each ablation. Lower MSE and higher SSIM are better.

Metric	No-CF	Horizon-only	Pruning-only
Retention MSE	**50** / 0 / 13	13 / **45** / 5	**49** / 0 / 14
MSE	**51** / 0 / 12	9 / **45** / 9	**47** / 0 / 16
SSIM	**53** / 0 / 10	8 / **46** / 9	**47** / 0 / 16

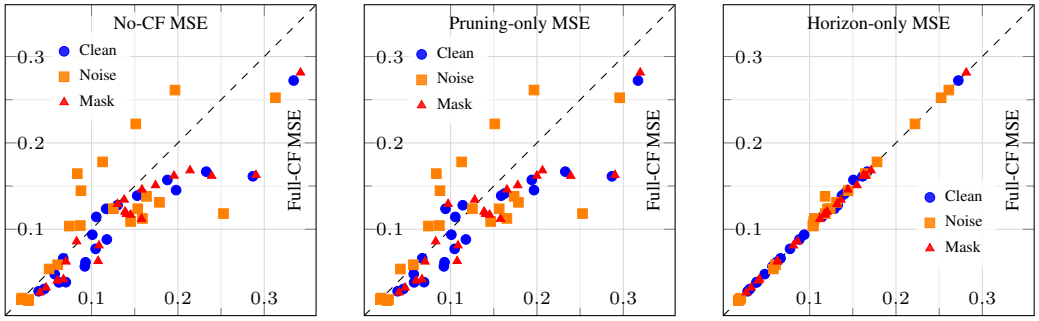
(a) Overall CF comparison

Compared with Full-CF	Lower MSE	Equal	Higher MSE
No-CF	**50**	0	13
Horizon-only	**13**	45	5
Pruning-only	**49**	0	14

(b) Retention MSE comparison

Figure 7 shows the individual online MSE comparisons underlying the aggregate results in Table 7a. The distance from the diagonal further indicates the magnitude of the performance difference between Full-CF and each ablation.

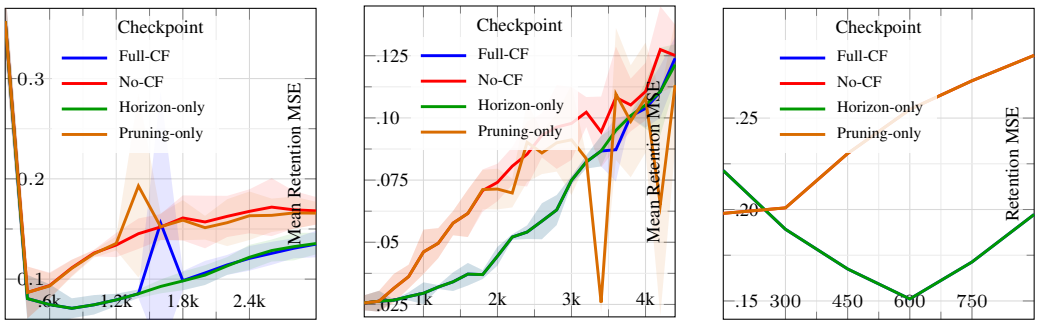
Against No-CF in Figure 8a, most points lie below the diagonal, with Full-CF achieving lower MSE in $51/63$ settings. A similar pattern is observed against Pruning-only in Figure 8b, where Full-CF achieves lower MSE in $47/63$ settings. In contrast, the Horizon-only results in Figure 8c are concentrated on or very close to the diagonal, with identical MSE in $45/63$ settings, showing that Horizon-only closely reproduces the performance of Full-CF.



(a) Full-CF vs No-CF **(b)** Full-CF vs Pruning-only **(c)** Full-CF vs Horizon-only

Figure 7: Pairwise online MSE comparison of Full-CF against (a) No-CF, (b) Pruning-only, and (c) Horizon-only across 63 dataset, memory-size, and cue settings. Clean, Noise, and Mask cues are shown by blue circles, orange squares, and red triangles, respectively. The dashed diagonal denotes equal performance; points below it indicate lower MSE for Full-CF.

Having established the relative effects of the two CF mechanisms from the aggregate results, we next examine whether these effects persist throughout continual updating rather than emerging only at the final evaluation. Figure 8 traces retention MSE across successive checkpoints using clean test cues and FFW readout, with grayscale datasets averaged at size 300, RGB datasets averaged at size 2000, and CIFAR10-DVS evaluated at size 150.



(a) Grayscale **(b)** RGB **(c)** CIFAR10-DVS

Figure 8: Retention MSE trajectories of early stored memories during continued online updates under Full-CF, No-CF, Horizon-only, and Pruning-only: (a) grayscale datasets averaged at size 300, (b) RGB datasets averaged at size 2000, and (c) CIFAR10-DVS at size 150. Results use clean test cues with EN representation and FFW readout; lower MSE is better.

The trajectories corroborate the aggregate results. Across all three settings, Full-CF closely tracks Horizon-only and generally maintains lower retention error than No-CF and Pruning-only as continual updating progresses. This separation is particularly clear for CIFAR10-DVS, where Full-CF and Horizon-only follow the same trajectory, while No-CF and Pruning-only exhibit substantially greater retention degradation. The temporal behaviour therefore supports the earlier finding that the adaptive horizon provides the primary protection against progressive overwrite, with pruning acting as a complementary safeguard against persistent trace dominance.

A.7.3 Temporal Retrieval and Readout Ablation Results

Table 8 extends the readout analysis in Section 5.4 with mean performance, sole-best counts, and modality-specific results. This separates the overall advantage of FFW from cases where alternative temporal readouts become competitive.

Table 8: Detailed temporal readout ablation over 63 online retrieval settings with Full-CF. Best counts report MSE / SSIM including ties. Δ_{LFW} denotes the mean MSE reduction relative to LFW.

	Mean Performance over all datasets					Best MSE / SSIM incl. ties			
Readout	MSE ↓	Δ_{LFW}	MSE sole best	SSIM ↑	SSIM sole best	Gray (36)	RGB (18)	DVS (9)	All (63)
FFW	**0.1119**	**5.9%**	**26**	**0.2763**	**36**	**26/32**	6/10	3/3	**35/45**
LFW	0.1189	0.0%	17	0.2245	9	3/0	**18/13**	**7/7**	28/20
DW	0.1149	3.4%	0	0.2390	1	3/0	6/7	**7/7**	16/14
WSW	0.1148	3.5%	4	0.2392	4	7/4	6/6	**7/7**	20/17

FFW achieves the lowest mean MSE (0.1119) and highest mean SSIM (0.2763), reducing mean MSE by 5.9% relative to LFW. It also records substantially more sole-best results, particularly for SSIM, showing that its overall advantage is not driven solely by ties. DW and WSW remain close in mean MSE but rarely provide uniquely best retrievals.

The modality breakdown reveals a different pattern. FFW dominates grayscale retrieval, whereas LFW is considerably more competitive on RGB data. For CIFAR10-DVS, LFW, DW, and WSW frequently tie, indicating little distinction among later temporal readouts for event-driven inputs. These results explain why FFW performs strongest overall while showing that the relative benefit of each readout depends on the input modality.

A.7.4 Comparison with Threshold Phasor Associative Memory

To evaluate SHONA against an established spike-based AM, we conduct a direct comparison with TPAM across the MNIST, EMNIST, and Fashion-MNIST datasets using identical stored patterns and retrieval cues.

For each dataset, two samples per class are selected, yielding 20 stored patterns for MNIST and Fashion-MNIST and 124 for EMNIST. Retrieval performance is assessed under consistent experimental conditions: clean cues, additive Gaussian noise ($\sigma = 0.5$), and 50% masking, with SHONA evaluated using the FFW readout. Table 9 summarizes the mean MSE, the difference between TPAM and SHONA, and the number of paired retrievals in which each model achieves the lower error.

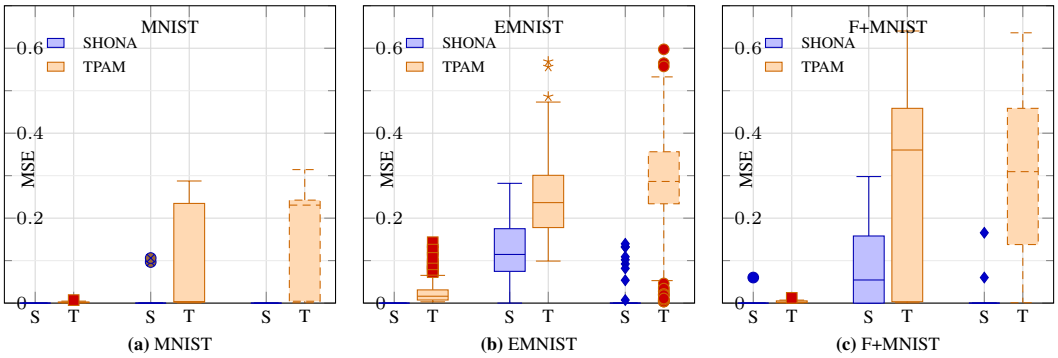
Table 9: Retrieval MSE of SHONA and TPAM under clean, Gaussian noise $\sigma = 0.5$, and 50% masked cues. Δ denotes the TPAM–SHONA MSE difference, with positive values indicating lower SHONA error; wins count paired retrievals with lower MSE.

Dataset	Cue	SHONA ↓	TPAM ↓	Δ	SHONA wins	TPAM wins
MNIST	Clean	**0.0000**	0.0021	+0.0021	**20**	0
	Noise	**0.0101**	0.1011	+0.0910	**19**	1
	Mask	**0.0000**	0.1729	+0.1729	**20**	0
EMNIST	Clean	**0.0000**	0.0252	+0.0252	**124**	0
	Noise	**0.1212**	0.2514	+0.1302	**112**	12
	Mask	**0.0069**	0.2713	+0.2645	**123**	1
F+MNIST	Clean	**0.0030**	0.0034	+0.0004	**19**	1
	Noise	**0.0860**	0.2813	+0.1953	**18**	2
	Mask	**0.0113**	0.2975	+0.2862	**20**	0

Across all three datasets, SHONA achieves the lower paired MSE in $475/492$ retrievals (96.5%), with the strongest separation appearing for corrupted cues. Under 50% masking, SHONA wins

163/164 retrievals while maintaining very low mean error across MNIST, EMNIST, and F+MNIST, whereas TPAM error rises substantially in all three cases. Gaussian noise is more challenging for SHONA, particularly on EMNIST, but it still achieves the lower error in most paired retrievals.

A further distinction appears between missing and perturbed information. SHONA is particularly tolerant to masking, most clearly on EMNIST, where the mean MSE remains at only 0.0069 despite half of the cue being removed. Although the strongest robustness is observed under masking, performance under Gaussian noise also remains comparatively strong. This indicates greater robustness to incomplete cues than to distributed perturbation, while the same recovery under masking is not observed for TPAM.



(a) MNIST **(b)** EMNIST **(c)** F+MNIST

Figure 9: Distribution of retrieval MSE for SHONA and TPAM on MNIST, EMNIST, and F+MNIST under clean, Gaussian-noise ($\sigma = 0.5$), and 50% masked cues. S and T denote SHONA and TPAM, respectively.

Figure 9 further shows that the advantage is not driven by isolated retrievals: under corrupted cues, the SHONA distributions remain substantially lower and more concentrated, particularly under masking, whereas TPAM exhibits broad high-error distributions. Notably, this separation is already evident at the 20-pattern scale used in TPAM's original image-retrieval experiments,[1] and remains evident when the comparison is extended to 124 stored patterns on EMNIST.

A.8 Robustness and Sensitivity Results

A.8.1 Cue Corruption Results

The robustness analysis in Section 5.5 is extended here by examining the effect of increasing corruption severity across all evaluated methods and by considering structural similarity in addition to MSE.

Table 10 shows that SHONA is particularly robust to increasing cue incompleteness. Under masking, SHONA achieves the lowest mean MSE at both 30% and 50%, with its error increasing by only 8.5% as half of the cue is removed. Although MHN exhibits a similarly small increase of 7.9%, its absolute error remains higher, with SHONA maintaining approximately 12% lower MSE at both masking levels. In contrast, SQHN, KNN, L_1, and Cosine deteriorate by 21.0%, 23.6%, 29.7%, and 17.5%, respectively.

The relative advantage of SHONA over these methods also becomes larger under stronger masking, its MSE advantage increases $(15.6\% \rightarrow 24.3\%)$ over SQHN, $(7.1\% \rightarrow 18.4\%)$ over KNN, $(7.1\% \rightarrow 22.3\%)$ over L_1, and $(9.3\% \rightarrow 16.3\%)$ over Cosine. This indicates that SHONA becomes increasingly competitive as the amount of available cue information is reduced.

Under Gaussian noise, SHONA remains competitive but exhibits a different sensitivity profile. Its mean MSE increases from 0.0950 at $\sigma = 0.3$ to 0.1178 at $\sigma = 0.5$, corresponding to an increase of 24.0%. SHONA nevertheless outperforms MHN at both noise levels, with 23.0% lower MSE at $\sigma = 0.3$ and 6.1% lower MSE at $\sigma = 0.5$. Although the direct similarity-based methods achieve lower MSE under Gaussian noise, SHONA already reaches a mean MSE below 0.1 at $\sigma = 0.3$, supporting effective retrieval under moderate perturbation.

[1] The original TPAM image-retrieval experiment stores $M = 20$ image patches of size $12 \times 12 \times 3$ and evaluates retrieval from noisy cues (Frady & Sommer, 2019).

Table 10: Robustness to increasing cue corruption under online retrieval. Values report mean MSE across the 21 dataset–capacity settings. Δ_{MSE} denotes the relative increase in MSE from the lower to the higher corruption level. Gain denotes SHONA's relative MSE reduction with respect to the method in the corresponding row; positive values favour SHONA. Lower MSE is better.

	Gaussian Noise					Masking				
	Mean MSE		Δ_{MSE}	SHONA Gain		Mean MSE		Δ_{MSE}	SHONA Gain	
Method	$\sigma = 0.3$	$\sigma = 0.5$		$\Delta_{.3}$	$\Delta_{.5}$	30%	50%		$\Delta_{30\%}$	$\Delta_{50\%}$
SHONA	0.0950	0.1178	24.0%	–	–	**0.1053**	**0.1142**	8.5%	–	–
MHN	0.1233	0.1254	1.7%	**+23.0%**	**+6.1%**	0.1197	0.1291	7.9%	+12.0%	+11.5%
SQHN	0.0689	0.0696	1.0%	−37.9%	−69.3%	0.1247	0.1509	21.0%	+15.6%	+24.3%
KNN	0.0636	0.0707	11.2%	−49.4%	−66.6%	0.1133	0.1400	23.6%	+7.1%	+18.4%
L_1	**0.0621**	**0.0631**	1.6%	−53.0%	−86.7%	0.1133	0.1470	29.7%	+7.1%	+22.3%
Cosine	0.0658	0.0731	11.1%	−44.4%	−61.1%	0.1161	0.1364	17.5%	+9.3%	+16.3%

Table 11 further examines whether these trends extend beyond pixel-wise reconstruction error. Under masking, SHONA achieves an SSIM of 0.3101 at 30% masking and the highest SSIM of 0.2747 at 50% masking. Thus, its robustness to increasingly incomplete cues is also reflected in the preservation of structural information. Under Gaussian noise, the direct similarity-based methods remain stronger in SSIM, consistent with the MSE results, while SHONA continues to outperform MHN at both noise levels.

Table 11: Structural similarity under increasing cue corruption for online retrieval. Values report mean SSIM across the 21 dataset–capacity settings. Higher is better.

	Gaussian Noise		Masking	
Method	$\sigma = 0.3 \uparrow$	$\sigma = 0.5 \uparrow$	$30\% \uparrow$	$50\% \uparrow$
SHONA	0.3134	0.2377	0.3101	**0.2747**
MHN	0.2189	0.2083	0.2231	0.1814
SQHN	**0.4168**	**0.4047**	0.2910	0.2081
KNN	0.3796	0.3308	0.3095	0.2279
L_1	0.4081	0.3951	**0.3198**	0.2242
Cosine	0.3929	0.3500	0.3013	0.2253

A.8.2 Temporal Horizon Results

The temporal-horizon analysis in Section 5.5 is further separated by cue condition and dataset to examine whether the overall sensitivity to T is driven by a particular form of input or data modality.

Table 12 shows that the preference for the shorter temporal horizon is consistent across Clean, Noise, and Mask cues. For all three conditions, $T = 300$ achieves both the lowest mean MSE and the highest mean SSIM. For Clean retrieval, mean MSE increases from 0.1390 at $T = 300$ to 0.1581 and 0.1730 at $T = 350$ and $T = 400$, respectively. The same trend occurs under Noise and Mask, showing that the overall result is not driven by any single cue condition.

Table 12: Online sensitivity to the retrieval horizon T under Full-CF. Values report mean performance over the representative-capacity settings across seven datasets. Lower MSE and higher SSIM are better.

	MSE ↓			SSIM ↑		
Cue	$T = 300$	$T = 350$	$T = 400$	$T = 300$	$T = 350$	$T = 400$
Clean	**0.1390**	0.1581	0.1730	**0.2430**	0.2281	0.2130
Noise	**0.1035**	0.1212	0.1334	**0.2225**	0.2177	0.1875
Mask	**0.1486**	0.1674	0.1834	**0.2174**	0.2066	0.1943

The dataset-level results in Table 13 further show that this behaviour is broadly consistent across grayscale, RGB, and event-based datasets. Across the seven datasets and three cue conditions, $T = 300$ achieves the lowest MSE in all 21 dataset–cue settings. The SSIM results follow the same trend in 18 of the 21 settings. The only exceptions are FMNIST under Noise and Mask and CIFAR10 under Noise, where $T = 350$ achieves slightly higher SSIM, although the differences are minimal.

These results show that SHONA operates effectively within a relatively short temporal window, with little benefit from extending the horizon. Despite $T = 350$ being commonly used as a practical

Table 13: Online sensitivity to the retrieval horizon T under Full-CF for each representative dataset–capacity setting. Lower MSE and higher SSIM are better.

		MSE ↓			SSIM ↑		
Dataset	Cue	$T = 300$	$T = 350$	$T = 400$	$T = 300$	$T = 350$	$T = 400$
MNIST	Clean	**0.1345**	0.1389	0.1429	**0.3375**	0.3248	0.3078
	Noise	**0.0931**	0.1090	0.1045	**0.3139**	0.3137	0.2480
	Mask	**0.1407**	0.1459	0.1489	**0.3004**	0.2880	0.2750
EMNIST	Clean	**0.1344**	0.1414	0.1527	**0.3425**	0.3241	0.2947
	Noise	**0.1104**	0.1237	0.1335	**0.2822**	0.2638	0.2290
	Mask	**0.1480**	0.1509	0.1632	**0.2970**	0.2707	0.2554
FMNIST	Clean	**0.1318**	0.1453	0.1535	**0.3170**	0.2962	0.2846
	Noise	**0.1020**	0.1125	0.1225	0.3279	**0.3526**	0.2923
	Mask	**0.1654**	0.1681	0.1850	0.2723	**0.2780**	0.2573
F+MNIST	Clean	**0.1431**	0.1571	0.1612	**0.3066**	0.2925	0.2882
	Noise	**0.1222**	0.1381	0.1392	**0.2269**	0.2230	0.2201
	Mask	**0.1490**	0.1619	0.1677	**0.2789**	0.2745	0.2734
CIFAR10	Clean	**0.0953**	0.1237	0.1400	**0.1424**	0.1343	0.1182
	Noise	**0.0448**	0.0540	0.0706	0.1358	**0.1378**	0.1131
	Mask	**0.0948**	0.1292	0.1497	**0.1376**	0.1259	0.1129
CIFAR100	Clean	**0.1044**	0.1278	0.1571	**0.1489**	0.1320	0.1164
	Noise	**0.0448**	0.0589	0.0760	**0.1530**	0.1275	0.1186
	Mask	**0.1064**	0.1342	0.1597	**0.1339**	0.1212	0.1072
CIFAR10DVS	Clean	**0.2298**	0.2722	0.3035	**0.1060**	0.0928	0.0814
	Noise	**0.2072**	0.2523	0.2873	**0.1175**	0.1055	0.0912
	Mask	**0.2358**	0.2814	0.3097	**0.1020**	0.0879	0.0789

operating range in prior spiking classification models (Na & Kim, 2025; Gygax & Zenke, 2025), SHONA demonstrates that a spiking AM model can achieve its best performance at $T \leq 350$ across cue conditions, datasets, and both evaluation metrics.

A.9 Model Pseudocode

Algorithm 1 SHONA online retrieval, update, and CF control

Input: cue $\boldsymbol{\xi}_{\text{in}}$, memory bank $\boldsymbol{\Xi} = \{\boldsymbol{\xi}_i\}_{i=1}^{N}$, horizon T, CF states $\boldsymbol{a}, \boldsymbol{s}$
Output: retrieved memory $\hat{\boldsymbol{\xi}}$ and updated memory state
Encode $\boldsymbol{\xi}_{\text{in}}$ into spike train $\boldsymbol{X} = \{\boldsymbol{x}(t)\}_{t=1}^{T}$ using Equation 21
Initialize $\boldsymbol{u} \leftarrow u_{\text{rest}}\mathbf{1}$, $\boldsymbol{\vartheta} \leftarrow \vartheta\mathbf{1}$, $\boldsymbol{c} \leftarrow \mathbf{0}$, $\mathcal{W} \leftarrow \varnothing$, $T_{\text{eff}} \leftarrow T$
for $t = 1$ **to** current T_{eff} **do**
  Update $\boldsymbol{u}(t)$ and $\boldsymbol{\vartheta}(t)$ using Equation 11
  Select winner $i^*(t)$ using Equation 12
  if $A(t) \neq \varnothing$ **then**
    Append $i^*(t)$ to $\mathcal{W}$ and $c_{i^*} \leftarrow c_{i^*} + 1$
    Apply post-firing dynamics using Equation 13
    Update $\boldsymbol{\xi}_{i^*}$ using Equations 14–15
    Recompute T_{eff} using Equations 16–17
  end if
end for
Read out $\hat{\boldsymbol{\xi}}$ from winner sequence $\mathcal{W}$ using the selected temporal readout in Appendix A.6
Update accumulated usage $\boldsymbol{a}$ using Equation 16 and streak state $\boldsymbol{s}$ using Equation 76
$i^\dagger \leftarrow \arg\max_i c_i$
if $a_{i^\dagger} > \tau_a$ **or** $s_{i^\dagger} > \tau_s$ **then**
  Reset $\boldsymbol{\xi}_{i^\dagger}$, $a_{i^\dagger}$, and $s_{i^\dagger}$
end if
return $\hat{\boldsymbol{\xi}}$, $\boldsymbol{\Xi}$, $\boldsymbol{a}$, $\boldsymbol{s}$

A.10 PARAMETERS

All SHONA experiments use a base spike horizon of $T = 350$ time steps. Periodic rate coding assigns each nonzero input component a firing rate within $[f_{\min}, f_{\max}]$, where $f_{\min} = 1$ and $f_{\max} = 50$. The spike-driven retrieval dynamics use a resting potential of $u_{\text{rest}} = -70$ and a baseline firing threshold of $\vartheta = -65$. The inhibitory and hyperpolarization potentials are $u_{\text{inh}} = -100$ and $u_{\text{hyp}} = -90$, respectively, with a refractory period of $\tau_{\text{ref}} = 15$ time steps. The membrane- and threshold-relaxation steps are $\eta_u = 0.8$ and $\eta_\vartheta = 0.4$, while the post-firing threshold increment is $\delta_\vartheta = 1$.

For winner-conditioned plasticity, the backward spike window is $P = 5$ and the overall update scale is $\sigma = 0.01$. The timing kernel uses $A_+ = 0.8$, $A_- = 0.3$, $\tau_+ = \tau_- = 5.0$, and $\rho = 0$. The value-dependent modulation is governed by $\xi_{\min} = 10^{-5}$, $\xi_{\max} = 1.0$, and $\mu = 0.9$.

For catastrophic-forgetting control, the adaptive-horizon gate uses $b = 1/3$, $K = 6$, and $n = 1$, with $\epsilon = 10^{-8}$ used in the normalized accumulated-usage update. The effective horizon is lower-bounded by one time step. A dominant memory is reset when its accumulated usage or consecutive-selection streak exceeds $\tau_a = 20$ or $\tau_s = 3$, respectively.

The memory capacities corresponding to the three experimental scales are $\{300, 1300, 2300\}$ for non-RGB datasets, $\{700, 2000, 3300\}$ for RGB datasets, and $\{150, 300, 500\}$ for CIFAR10-DVS.