

Emergent self-attention from astrocyte-gated associative memory dynamics

Arnaud Vivet^{1,*} and Alex Arenas^{1,2,†}

¹*Departament d'Enginyeria Informàtica i Matemàtiques, Universitat Rovira i Virgili, 43007 Tarragona, Spain*

²*Complexity Science Hub Vienna, Metternichgasse 8, 1030 Vienna, Austria*



(Received 1 March 2026; accepted 19 May 2026; published 11 June 2026)

We introduce a Hopfield-type associative memory in which effective connectivity is multiplicatively modulated by astrocytic gains evolving under an entropy-regularized replicator equation. The coupled neuron-astrocyte dynamics admit a Lyapunov function, ensuring global convergence. At fixed points, astrocytic gains implement a softmax-normalized allocation over pattern similarity scores, yielding a mechanistic realization of self-attention as emergent routing on the gain simplex. In regimes of high memory load and interference, the model significantly improves retrieval accuracy relative to classical Hopfield dynamics and recent neuron-astrocyte baselines. These results establish a dynamical systems framework linking glial modulation, competitive resource allocation, and attentionlike computation.

DOI: [10.1103/6q3q-vfs9](https://doi.org/10.1103/6q3q-vfs9)

I. INTRODUCTION

Hopfield networks revolutionized our understanding of how collective neuronal dynamics implement associative memory with biological plausibility [1]. In their original formulation, neurons evolve under Hebbian connectivity that stores a finite set of patterns as energy minima. Although this framework provided a foundational bridge between statistical physics and neural computation, its practical applicability was limited by low storage capacity and the proliferation of spurious attractors under high pattern load or correlation [2].

Interest resurged with generalized energy-based formulations, dense associative memory and modern Hopfield networks, that replace the quadratic Hebbian energy with higher-order or exponential interaction terms, dramatically expanding capacity and retrieval stability [3,4]. A key insight from this line of work is that the fixed-point readout of modern Hopfield networks is mathematically equivalent to the scaled dot-product attention mechanism at the core of transformer architectures [5,6]. This correspondence has reframed associative memory dynamics as a candidate substrate for attention and contextual reasoning in both artificial and biological systems.

Neuroscience has concurrently revealed that the brain's computational resources extend well beyond neurons. Astrocytes, glial cells long considered purely supportive, actively regulate synaptic transmission, integrate signals across local microcircuits, and contribute to memory-related plasticity [7–12]. Through Ca^{2+} signaling, neurotransmitter uptake, and connexin-mediated gap-junction coupling, astrocytes modu-

late neuronal excitability and impose structured heterogeneity in effective synaptic strengths [9,11,12]. Astrocytes have also been implicated in controlling circuit-level dynamical regimes, including the modulation of gamma-band synchronization associated with memory processing [13–15]. Critically, astrocytes are now recognized as active participants in memory encoding and consolidation: Recent experimental work demonstrates that astrocyte ensembles selectively encode and stabilize salient experiences, with individual astrocytes reactivated during memory retrieval undergoing noradrenaline-dependent molecular tagging that promotes consolidation of specific traces [16,17]. These findings extend the classical engram framework [18] by assigning astrocytes a mechanistic role in the selection and stabilization of memory representations, motivating their inclusion in computational models of associative memory.

Building on this foundation, recent theoretical work has proposed that astrocytic modulation could implement attentionlike computation by providing a flexible, content-dependent reweighting of neuronal ensembles during retrieval [19,20]. In this view, self-attention corresponds to a gain reallocation across stored patterns driven by their relevance to the current query state. However, a precise dynamical account of how softmax-normalized gains arise as an emergent, self-organized property of neuron-astrocyte interaction, rather than being imposed by architectural design, has remained absent.

Here, we close this gap by introducing a Hopfield-type associative memory in which effective connectivity is multiplicatively modulated by astrocytic gains that evolve on the probability simplex under an entropy-regularized replicator equation. The model does not aim to reproduce a specific microscopic astrocytic biochemical pathway, but instead treats astrocytic modulation as a coarse-grained control variable acting on memory-related synaptic pathways. The coupled neuron-astrocyte system admits a Lyapunov function, ensuring convergence to stationary points (equilibria) at which gains implement a Gibbs-Boltzmann (softmax) allocation over pattern similarity scores. The stationary readout is thus

*Contact author: arnau.vivet@urv.cat

†Contact author: alexandre.arenas@urv.cat

Published by the American Physical Society under the terms of the Creative Commons Attribution 4.0 International license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

a convex combination of stored patterns with self-attention weights—emerging purely from the competitive resource dynamics of glial modulation, without any explicit attention mechanism being prescribed. Under high memory load and pattern interference, this mechanism substantially improves retrieval accuracy relative to classical Hopfield dynamics and recent neuron-astrocyte baselines, establishing a dynamical systems framework that links glial gain control, competitive resource allocation, and the emergence of attentionlike computation in neural circuits.

II. ASTROCYTE-GATED ASSOCIATIVE MEMORY

To ground the proposed astrocyte-neuron model, we begin by recalling the continuous-time classical Hopfield network, the minimal framework for associative memory. It is a class of recurrent neural network in which we consider a population of rate-based neurons evolving under a fully connected synaptic matrix. One of the key properties of this framework is that it can be derived from a Lyapunov function that guarantees convergence, enabling a precise characterization of memory retrieval as well as interference and capacity limits [1,2]. For these reasons, the Hopfield model provides a natural substrate on which biologically motivated extensions can be constructed. More precisely, we consider N rate units $x(t) \in \mathbb{R}^N$ with elementwise activation function $\phi(x) = \tanh(\sigma x) \in (-1, 1)^N$ and synaptic weights $W \in \mathbb{R}^{N \times N}$:

$$\tau_x \dot{x}^i = -x^i + \sum_j W^{ij} \phi(x^j), \quad (1)$$

where τ_x is the neuronal rate time constant. It sets the intrinsic timescale over which the firing-rate state $x(t)$ relaxes toward the drive $W \phi(x)$ against the leak term. The synaptic matrix W is constructed using Hebbian storage of K binary patterns $\xi_\mu \in \{-1, 1\}^N$ located at the vertices of the hypercube:

$$W_H = \frac{1}{N} \sum_{\mu=1}^K \xi_\mu^i \xi_\mu^j. \quad (2)$$

Alternatively, in matrix form we have $W_H = \frac{1}{N} \Xi \Xi^T$, where $\Xi \equiv (\xi_1, \dots, \xi_K) \in \{-1, 1\}^{N \times K}$ is a binary matrix where patterns are stored as columns. Throughout the paper, we use superscripts to label neurons and subscripts to label memory patterns.

A. Astrocyte-gated retrieval as patternwise gain control

We extend the classical Hopfield rate dynamics by introducing K astrocytic gain variables $p = (p_1, \dots, p_K)$ that modulate the contribution of each stored pattern to the synaptic matrix. The gains are constrained to the probability simplex,

$$p_\mu \geq 0 \ (\forall \mu), \quad \sum_{\mu=1}^K p_\mu = 1, \quad (3)$$

so modulation is non-negative and globally conserved. Note that the p variables are a phenomenological pathway-level gain and not a directly measurable astrocytic state, and that the simplex constraint and entropic term as the minimal represen-

tation of finite shared modulatory resources and regularized competitive allocation.

The resulting neuronal dynamics read

$$\tau_x \dot{x}^i = -x^i + \sum_{j=1}^N W^{ij}(p) \phi(x^j), \quad (4)$$

with effective synaptic matrix

$$W(p) \equiv \frac{K}{N} \sum_{\mu=1}^K p_\mu \xi_\mu \xi_\mu^T = \frac{K}{N} \Xi \text{diag}(p) \Xi^T. \quad (5)$$

Thus, p_μ acts as a multiplicative gain on the μ th Hebbian outer product. The classical Hopfield coupling is recovered for uniform gains $p_\mu = 1/K$, for which $W(p) = W_H$.

We model astrocytic modulation as an adaptive allocation process governed by a softmax-regularized replicator flow (hereafter “s-replicator”), which is the same flow described in Ref. [21] under the name “adjusted replicator equation” in the context of learning in games.

$$\tau_p \dot{p}_\mu = p_\mu \left(F_\mu(x, p_\mu) - \sum_{v=1}^K p_v F_v(x, p_v) \right), \quad (6)$$

which preserves the simplex constraints. The fitness is defined as

$$F_\mu(x, p_\mu) \equiv f_\mu(x) - T \ln p_\mu, \quad T > 0, \quad (7)$$

where

$$f_\mu(x) \equiv \frac{1}{2N} \left(\sum_{j=1}^N \xi_\mu^j \phi(x^j) \right)^2 \quad (8)$$

is the squared overlap with pattern μ and the logarithmic term acts as an entropic (temperaturelike) regularizer that discourages collapse of the gains onto a single pattern. Throughout, we use the same neuronal nonlinearity as in the classical model, $\phi(x^j) = \tanh(\sigma x^j)$ with $\sigma > 0$.

Note that at an interior fixed point of Eq. (6), the fitnesses are equal:

$$F_\mu(x, p_\mu) = \sum_v p_v F_v(x, p), \quad \forall \mu, \ p_\mu > 0.$$

This means that $f_\mu(x) - T \ln p_\mu$ is constant for all μ . However, since some overlaps $f_\mu(x)$ are larger than others, p_μ compensates as

$$p_\mu^* = \frac{\exp(f_\mu(x)/T)}{\sum_v \exp(f_v(x)/T)}.$$

See Appendix A 1 a for more details of this derivation.

We study retrieval dynamics by initializing the neuronal state $x(0)$ as a corrupted version of one of the stored patterns. Unless stated otherwise, the astrocytic gains are initialized uniformly, $p_\mu(0) = 1/K$ for all μ . During the dynamics, Eq. (6) reallocates gain toward patterns with larger instantaneous overlap $f_\mu(x)$, thereby reshaping the effective coupling $W(p)$ in Eq. (5). This positive feedback selectively amplifies the synaptic contribution of the most compatible patterns and suppresses competing ones, reducing interference and effectively increasing the basin of attraction of the target memory.

The biophysical motivation for Eqs. (4) and (6) is the following. We view each stored pattern as an engramlake neuronal ensemble and interpret the gain p_μ as an effective proxy for astrocyte-mediated modulation of the synaptic pathways supporting that ensemble. Specifically, p_μ summarizes (at a coarse-grained level) how local astrocytic activity can bias presynaptic release probability and thus rescale the effective strength of the synapses recruited by pattern μ . The factor p_μ should therefore not be read as a literal cellular mechanism by which astrocytes identify a global memory label μ . Rather, the microscopic picture is local: Nearby astrocytic processes modulate synapse-level gains on coactive pathways. Synapses repeatedly coactive during encoding define, in the coarse-grained Hopfield description, an engramlake subnetwork represented by the Hebbian outer product $\xi_\mu \xi_\mu^\top$. The variable p_μ is the average effective modulatory gain of this subnetwork. Replicatorlike competition then arises only after these local or pathway-level gains are normalized under a finite modulatory budget.

This abstraction is inspired by evidence that synaptic activity can evoke highly localized astrocytic Ca^{2+} signals in perisynaptic microdomains adjacent to individual synapses [22]. In several experimental settings, astrocytic Ca^{2+} elevations have been reported to modulate presynaptic release probability and thereby reshape effective synaptic transmission, consistent with an activity-dependent astrocytic gating of synaptic pathways [23]. At the same time, the extent to which Ca^{2+} -dependent astrocytic signaling modulates excitatory transmission and plasticity under baseline conditions remains debated and depends on preparation and stimulation regime [24]; accordingly, we treat p_μ as a phenomenological variable that aggregates multiple microscopic mechanisms rather than as a direct readout of any single molecular process.

Finally, we constrain the astrocytic gains to the simplex, $p \in \Delta^{K-1}$, i.e., $p_\mu \geq 0$ and $\sum_{\mu=1}^K p_\mu = 1$. In the model, this implements a finite modulatory capacity: Increasing gain for one pattern necessarily reduces the gain available to others, thereby enforcing competitive allocation across pathways. This is a deliberate modeling device that operationalizes resource limitation and homeostatic competition at the level of pattern gains, rather than a claim that the brain explicitly maintains a globally conserved scalar “budget” [25].

The neuron-astrocyte coupling is driven by a pattern-match functional $f_\mu(x)$, which quantifies how strongly the current neuronal state expresses pattern μ (via its overlap with ξ_μ). We map this match to a fitness $F_\mu(x, p_\mu)$ that controls the gain dynamics: patterns with larger $F_\mu(x, p_\mu)$ are preferentially upweighted by Eq. (6). The additional logarithmic term in $F_\mu(x, p_\mu)$ acts as an entropic regularizer, penalizing highly concentrated allocations and preventing winner-take-all behavior; consequently, the stationary distribution of gains is softmaxlike, with sharpness set by the parameter T . Note that the term stands for an entropylike homeostatic feedback or diminishing-returns effect that opposes allocating nearly all modulatory gain to one pathway; other convex saturation mechanisms could produce related soft competition, but the logarithmic choice yields a closed-form Gibbs-Boltzmann fixed point and an analytically transparent Lyapunov structure.

B. Analytic results

We now summarize three analytic results that organize the theoretical picture and will be used throughout the remainder of the paper. First, we show that the coupled neuron-astrocyte dynamics admit a global Lyapunov function, so trajectories are dissipative and converge to the set of stationary points rather than exhibiting sustained oscillations or chaos. This structural result justifies focusing on equilibria. Second, we characterize these equilibria explicitly via coupled fixed-point equations for (x^*, p^*) , which reveal how astrocytic modulation implements a temperature-controlled, softmaxlike allocation over patterns at stationarity. Finally, we connect the framework back to the classical Hopfield model by identifying limiting regimes in which the gain distribution becomes uniform and the effective synaptic matrix reduces to W_H . Together, these three results provide (1) a convergence guarantee, (2) an interpretable description of the attractors, and (3) a consistency link to the standard associative memory baseline.

1. The system is a global gradient flow

A key property of the classical Hopfield network is the existence of an energy (Lyapunov) function that decreases along the trajectories, thus guaranteeing convergence of the dynamics. Recently, Ref. [26] showed that the Hopfield model is actually a gradient descent under a specified metric (see Appendix A 1 b). Being a gradient descent is a stronger geometric statement than being Lyapunov, and in that sense, our construction extends this geometric structure to the product of neuronal states and simplex-valued gains. As shown in Appendix A 1, our system can also be written as a gradient flow whose energy is given by the following scalar quantity:

$$\mathcal{L}(x, p) = KT \sum_{\mu} p_{\mu} \ln \frac{p_{\mu}}{\pi_{\mu}} - KT \ln Z + \sum_i^N x^i \phi(x^i) - L(x^i), \quad (9)$$

where $\pi_{\mu} = \frac{1}{Z} e^{f_{\mu}(x)/T}$, with $Z = \sum_{\mu} e^{f_{\mu}(x)/T}$ and $L(x^i) = \frac{1}{\sigma} \ln \cosh(\sigma x^i)$. The gradient-flow condition implies that this quantity decreases along the trajectories:

$$\frac{d}{dt} \mathcal{L}(x, p; t) = -\|\nabla \mathcal{L}(x, p; t)\|^2 \leq 0.$$

The potential is composed of a gain-dependent part that can be written as a temperature-scaled KL (Kullback-Leibler) divergence from the Gibbs distribution plus the corresponding log-sum-exp free energy. Expanding the first term, we get $\mathcal{L}(x, p) = -K \langle p, f \rangle + KT \langle p, \ln p \rangle + [\langle x, \phi \rangle - L(x)]$, which also identifies $-K \langle p, f(x) \rangle$ as the gain-weighted Hebbian interaction, $KT \langle p, \ln p \rangle$ as the entropic cost, and $\langle x, \phi(x) \rangle - L(x)$ as the classical continuous-Hopfield activation contribution. Since $\mathcal{L}$ is nonincreasing and constant only at stationary points, the long-time behavior is fully determined by the coupled equilibria (x^*, p^*) , which we analyze next.

2. Fixed-point analysis

The equilibrium (x^*, p^*) satisfies the coupled fixed-point condition

$$x^* = W(p^*)\phi(x^*), \quad (10)$$

$$p_\mu^* = \frac{\exp(f_\mu(x^*)/T)}{\sum_v \exp(f_v(x^*)/T)} = \text{softmax}_\mu\left(\frac{f_\mu(x^*)}{T}\right). \quad (11)$$

This way, initializing the neuron states as a corrupted version of one of the stored memories ξ_k , we say that retrieval is successful when the activated state $\phi(x^*)$ has large overlap with the target pattern ξ_k . In the next section, we define the retrieval error, as a measure of the degree to which the fixed point corresponds to the intended pattern.

This can be interpreted as a mixture of (rank 1) experts [27] commonly expressed as $x^* = \sum_\mu g_\mu(x^*)E_\mu(\phi(x^*))$, where the self-attention-like routing $g_\mu(x) \equiv p_\mu$ emerges naturally from the s-replicator and the experts are rank-1 matrices $E_\mu(\phi(x)) \equiv \xi_\mu \xi_\mu^\top \phi(x)$. In this interpretation, astrocytic modulation provides the routing mechanism that selects the relevant memory patterns in a context-dependent and dynamical manner.

These equilibrium relations make clear that the model reduces to the classical Hopfield network whenever the gain distribution is forced to remain (or becomes) uniform, which we discuss next through two limiting regimes.

3. Recovering the classical Hopfield model

Our model recovers the classical associative memory in two dynamical regimes. For the first case, we take the gain timescale to be infinite $\tau_p \rightarrow \infty$ such that the astrocytic influence remains constant for all time $\dot{p}_\mu = 0$ (see Appendix A 4). Since the modulation is initialized uniformly $p_\mu = 1/K$, we recover the classical Hopfield mechanism

$$\tau_x \dot{x} = -x + W(1/K)\phi(x). \quad (12)$$

This should be interpreted as the astrocytic processes being frozen, in which case there may not be a modulatory influence on the neuron dynamics. The second limit in which we recover the classical model is when we take $T \rightarrow \infty$, in which case the entropic term dominates and the only stable gain distribution is again uniform $p_\mu = 1/K$. In fact, if we divide the astrocytic component of Eq. (6) by T , everything is zero except for the logarithmic terms where we get $\ln p_\mu = \langle \ln p \rangle$, forcing $p_\mu = 1/K$ for all μ ; since the gains are initialized at $p_\mu(0) = 1/K$, they stay there for all time (see Appendix A 4).

C. Simulation results

Our focus is on how two control knobs shape retrieval: (1) the relative timescales of neuronal relaxation and astrocytic routing, quantified by τ_x and τ_p , and (2) the selectivity parameter T , which sets the strength of entropic regularization in the gain dynamics. Because our dynamics are a gradient flow on a potential (Sec. II B), their convergence is guaranteed in principle; numerically, however, we observe a competition between intrinsic convergence time and the finite simulation horizon

t_f . We therefore report both end-point observables and empirical convergence times to disentangle genuine failure from slow convergence. We emphasize that these simulations are intended as proof-of-principle demonstrations of dynamical regimes controlled by (τ_x, τ_p, T) .

1. Dynamical analysis

In this first analysis, the goal is to understand the long-term behavior of the system. To that end, we define two observables, one representative of the neuron component and the other for the astrocytic gains. Our focus is on how two control knobs shape retrieval: (1) the relative timescales of neuronal relaxation and astrocytic routing, quantified by τ_x and τ_p , and (2) the selectivity parameter T , which sets the strength of entropic regularization in the gain dynamics. Because our dynamics are Lyapunov (Sec. II B), convergence is guaranteed in principle; numerically, however, we observe a competition between intrinsic convergence time and the finite simulation horizon t_f . We therefore report both end-point observables and empirical convergence times to disentangle genuine failure from slow convergence.

We simulate a network of $N = 30$ neurons storing $K = 100$ binary (overlapping) patterns $\{\xi_\mu\}$ via Eq. (5). To probe retrieval, we corrupt a reference memory $\xi_0 \rightarrow \xi_0^\eta$ by flipping n randomly chosen entries, defining the noise level $\eta := n/N = 0.2$ (six flipped bits). We initialize the neuronal state with the corrupted memory, $x(t_i) = \xi_0^\eta$ (see Appendix A 5).

We quantify neuronal retrieval at t_f using a *soft* Hamming error,

$$\text{Error} = \epsilon(t_f) := \frac{1}{2} \sum_{i=1}^N |\xi_0^i - \phi(x^i(t_f))|, \quad (13)$$

which reduces to the standard Hamming distance when $\phi(x^i) \in \{-1, 1\}$ and satisfies $\epsilon_{\text{soft}} \in [0, N]$.

To quantify how concentrated the gain vector $p(t_f)$ is over memories, we compute its Shannon entropy $H[p]$ and the associated perplexity

$$P := \exp(H[p]) = \exp\left(-\sum_{\mu=1}^K p_\mu \ln p_\mu\right). \quad (14)$$

The reason for computing the perplexity instead of the entropy is that it converts the entropy into an effective number of active memory components. In particular, it satisfies $P \in [1, K]$, with $P = 1$ indicating near winner-take-all routing and $P \approx K$ indicating near-uniform allocation. Unless stated otherwise, we set $\tau_x = 1$, $\tau_p = 1$, and $T = 0.01$, and vary one parameter at a time.

Varying astrocyte timescale τ_p (keeping $\tau_x = 1$): Mechanistically, τ_p controls how rapidly the routing weights $p_\mu(t)$ track the instantaneous overlaps $f_\mu(x(t))$; small τ_p yields fast reweighting of $W(p)$ toward the correct pattern, whereas large τ_p delays this reshaping; i.e., τ_p parameter controls the gain $p(t)$ response time (Fig. 1). If we let astrocytic modulation be very fast ($\tau_p \rightarrow 0$), the gains quickly concentrate on the most compatible memories, so the retrieval is fast and we get both low error and low perplexity. As τ_p increases, we see the gain adaptation become slower and the system needs more time to route the right memory. With a fixed simulation

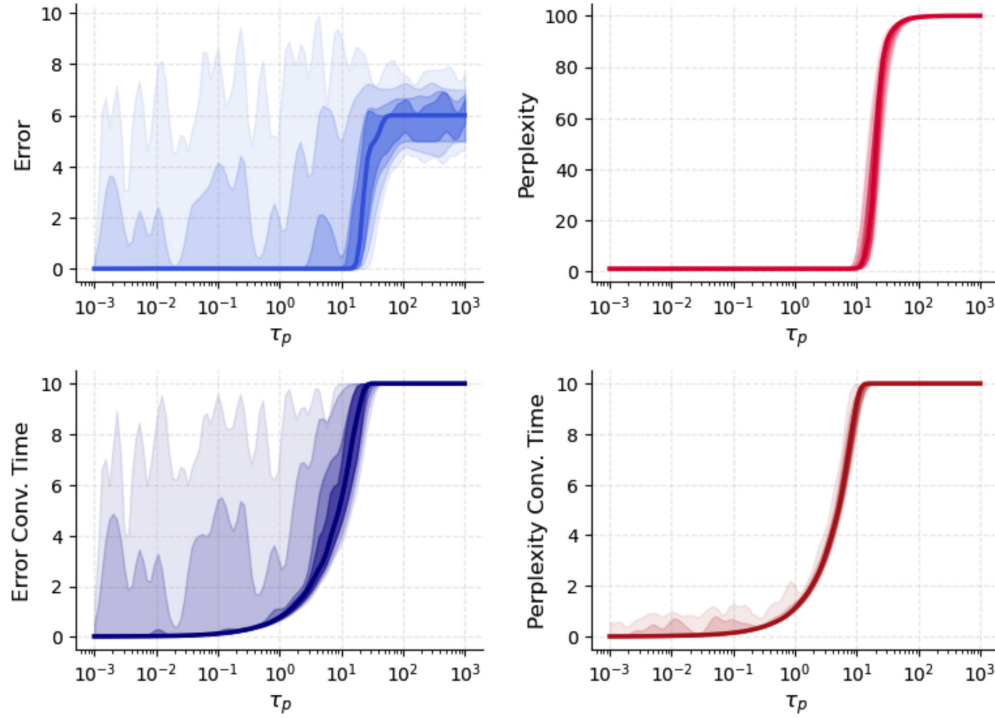


FIG. 1. End-of-run retrieval error $\epsilon(t_f)$ and gain perplexity $P(t_f)$ as functions of the astrocytic timescale τ_p (with $\tau_x = 1$), together with the corresponding convergence times. Solid lines show medians across trials; shaded bands denote percentile ranges (5, 95), (10, 90), (20, 80), and (25, 75).

time, it looks like performance gets worse beyond $\tau_p \sim 10$, but this behavior is explained simply because the simulation is stopped before convergence. In the extreme $\tau_p \rightarrow \infty$ case, since the modulation essentially becomes frozen, the classical Hopfield regime is recovered, which performs poorly in this high memory storage regime (see also Appendix A 4).

Varying neuron timescale τ_x (keeping $\tau_p = 1$): τ_x controls how fast neurons $x(t)$ relax (Fig. 2). Setting $\tau_x \rightarrow 0$, the neuron relaxation is essentially instantaneous; this makes the neuron configuration “commit” to an attractor too early, before the modulatory signal has time to reshape the landscape. Then, the gain modifies the effective connectivity around the wrong attractor, leading to high confidence (low perplexity) on the wrong memory. On the other hand, as we increase τ_x , we see a decrease in the error as well as an increase in perplexity. High perplexity here means that multiple patterns share similar overlap, so routing stays diffuse. This seemingly paradoxical region happens because as the neuron dynamics become slower, they do not have enough time to break the symmetry between patterns with similar fitness (see Appendix A 3). In fact, if we increase the time limit, symmetry breaks and we get high pattern selectivity again. If we keep increasing to $\tau_x \rightarrow \infty$, the convergence time increases superlinearly and we see the error stay exactly at $\epsilon(t_f) = 6$ (in our simulations, this happens at $\tau_x \sim 10^2$, simply because the simulation time is bounded) (see Appendix A 4).

Varying temperature T (keeping $\tau_p = \tau_x = 1$): Thus, T tunes a selectivity-robustness trade-off: Low T yields sharp routing (small P) and effective interference suppression, whereas high T enforces near-uniform gains and recovers Hopfield-like behavior; i.e., temperature controls astrocyte

selectivity (Fig. 3). When $T \rightarrow 0$, both error and perplexity are low since the regularization effect is weak and becomes highly selective. When $T \rightarrow \infty$, the error increases because the entropy dominates, meaning that p stays close to uniform and once again we recover the Hopfield regime (see Appendix A 4).

2. Comparative retrieval performance

We benchmark retrieval performance against (1) the classical Hopfield network and (2) the neuron-astrocyte associative-memory model of Kozachkov *et al.* [20] (see Fig. 4). To enable a dense scan over memory load and corruption, we set $N = 20$ and vary the number of stored random binary patterns from $K = 2$ to $K = 200$. For each (K, n) condition, we generate an independent pattern matrix Ξ , corrupt the target pattern ξ_0 by flipping n randomly chosen entries (equivalently, corruption fraction $\eta = n/N$), and initialize the network as $x(t_i) = \xi_0^{(n)}$ (simulation protocol and model-specific parameters are reported in Appendix A 5). All models are evaluated under the same initialization and the same finite simulation horizon t_f .

We quantify retrieval at time t_f by binarizing the final state and computing the Hamming error with respect to the target:

$$\epsilon(t_f) = \frac{1}{2} \sum_{i=1}^N |\xi_0^i - \text{sign}(x^i(t_f))|, \quad (15)$$

so that $\epsilon \in [0, N]$ and $\epsilon = 0$ indicates perfect retrieval. For statistical stability, we repeat each condition 50 times with independently sampled Ξ and report the mean error.

Across the tested (K, n) grid, our model yields lower mean retrieval error than both baselines, with the largest gains ap-

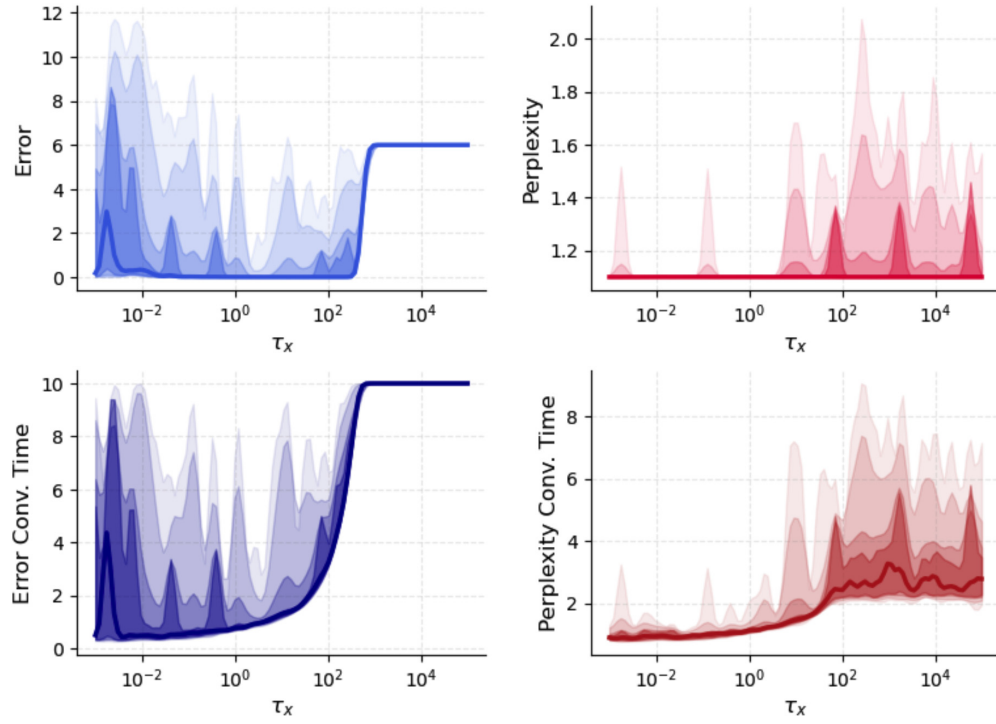


FIG. 2. Final-time retrieval error $\epsilon(t_f)$ and gain perplexity $P(t_f)$ as functions of the neuronal timescale τ_x (with $\tau_p = 1$), together with the corresponding convergence times. Solid lines show medians across trials; shaded bands denote percentile ranges (5, 95), (10, 90), (20, 80), and (25, 75).

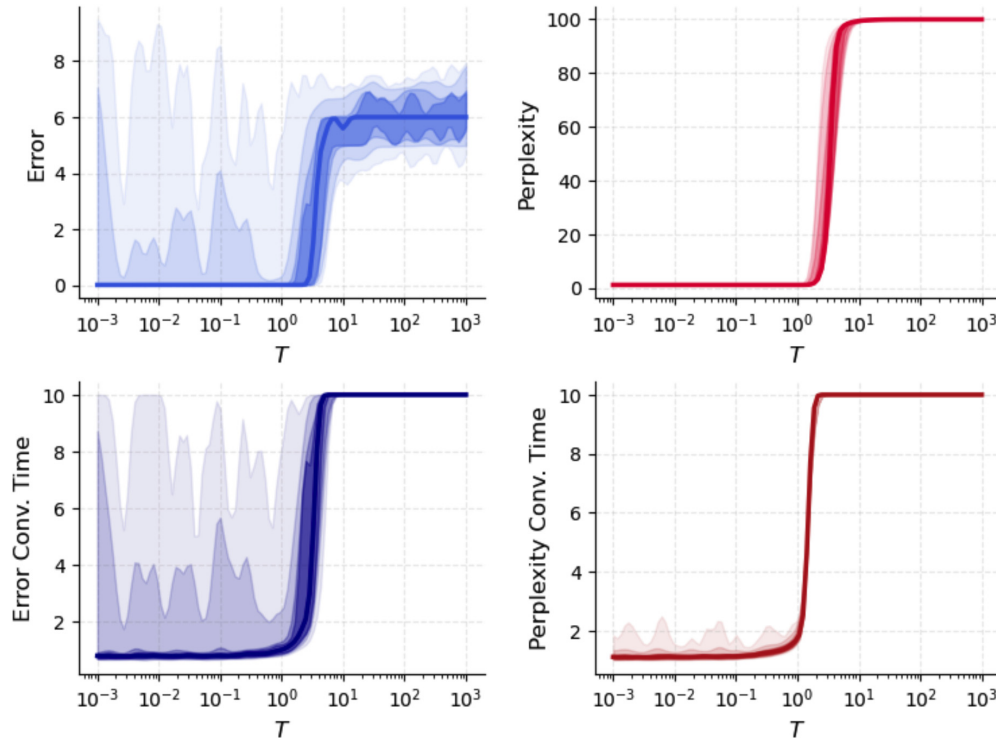


FIG. 3. Final-time retrieval error $\epsilon(t_f)$ and gain perplexity $P(t_f)$ as functions of temperature T (with $\tau_x = \tau_p = 1$), together with the corresponding convergence times. Solid lines show medians across trials; shaded bands denote percentile ranges (5, 95), (10, 90), (20, 80), and (25, 75).

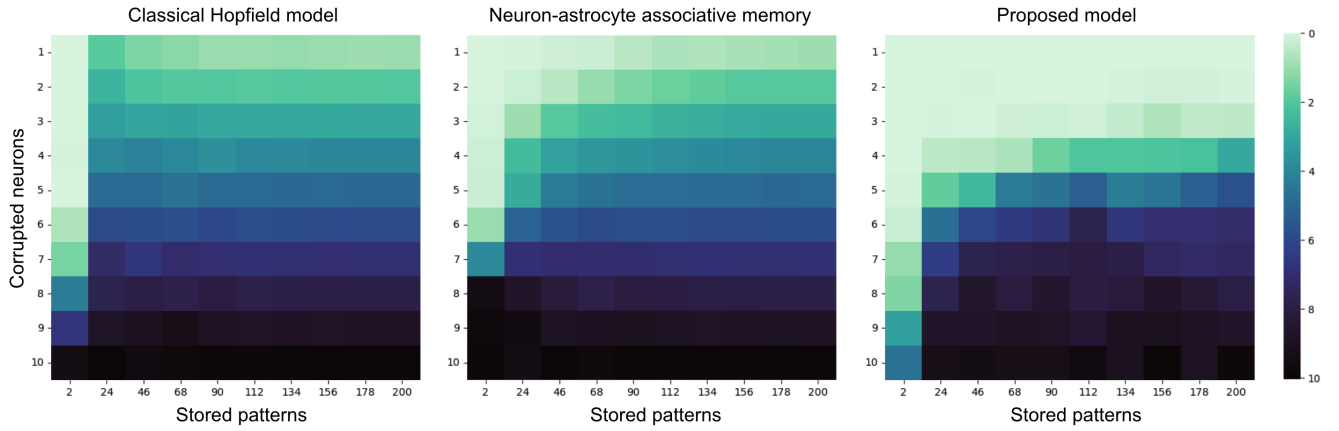


FIG. 4. Retrieval benchmark across models. Each heat map reports the mean Hamming retrieval error $\epsilon(t_f)$ (lighter indicates lower error) as a function of memory load K (x axis) and corruption level n (y axis; number of flipped bits in the query pattern). Each entry is averaged over 50 independent random pattern realizations.

pearing at high memory loads where interference is strongest (Fig. 4).

III. DISCUSSION

We introduced an astrocyte-gated extension of a Hopfield-type associative memory in which astrocytic processes allocate a finite modulatory capacity across stored patterns. By dynamically reweighting pattern-specific contributions to the effective connectivity, this additional degree of freedom reduces interference during retrieval and enlarges the regime in which accurate recall is achieved at high memory load.

We interpret each stored pattern as a memory-specific (engramlike) neuronal ensemble and the gain variable p_μ as a coarse-grained proxy for the effective strength of astrocyte-mediated modulation of the synaptic pathway supporting pattern μ . This abstraction is motivated by the tripartite-synapse framework, in which astrocytes integrate local activity and can modulate synaptic efficacy through Ca^{2+} -dependent signaling and related pathways [7–12]. Importantly, we do not interpret p_μ as a direct readout of any single molecular mechanism; rather, it summarizes pathway-level modulation at the scale relevant for associative retrieval.

A central modeling assumption is the simplex constraint $\sum_\mu p_\mu = 1$, which enforces competitive allocation: Increasing gain for one pathway necessarily reduces gain available to others. This operationalizes finite modulatory capacity (e.g., limited signaling/metabolic resources distributed across microdomains) and homeostatic competition at the level of pattern gains, without implying that the brain literally implements a globally conserved scalar “budget.”

Mechanistically, the gain dynamics implement state-dependent competition. When the current neuronal state is more compatible with pattern μ , the corresponding match $f_\mu(x)$ increases the fitness $F_\mu(x, p_\mu)$ and thereby upweights p_μ , which amplifies the μ th contribution to $W(p)$ while suppressing competitors. The logarithmic term in $F_\mu(x, p_\mu)$ acts as an entropic regularizer: It penalizes highly concentrated allocations and discourages winner-take-all routing. In this view, the temperature parameter T controls selectivity: Low T yields sharp, concentrated routing, whereas high T enforces

broader, near-uniform modulation and recovers Hopfield-like behavior. The joint neuron-astrocyte system remains Lyapunov consistent (it admits a global Lyapunov function), ensuring convergence of the coupled dynamics.

Modern Hopfield networks and related dense-memory models achieve softmaxlike retrieval through neuronal energy descent shaped by higher-order or exponential storage functions [3,4,6,19,20]. Our model produces a similar normalized reweighting, but via a different mechanism: Normalization arises from competitive dynamics on the gain simplex (replicator-type routing) rather than being hardwired into the neuronal energy landscape. In regimes where gains adapt rapidly, the stationary allocation approaches

$$p_\mu^* \propto \exp\left(\frac{f_\mu}{T}\right),$$

so retrieval can be viewed as operating with a dynamically reweighted subset of memories. Compared to the classical Hopfield model, where high load is associated with numerous spurious mixture states, our routing variable tends to concentrate weight on a small set of candidates under ambiguity, consistent with the low-perplexity regimes observed in simulations.

In addition, the degree to which attentionlike structure is explicit in our framework depends on the choice of pattern score. While the main model uses a squared-overlap score $f_\mu(x) \propto \langle \xi_\mu, \phi(x) \rangle^2$ [hence invariant under $\langle \xi_\mu, \phi(x) \rangle \mapsto -\langle \xi_\mu, \phi(x) \rangle$], one may alternatively define a linear-overlap score $f_\mu(x) = \langle \xi_\mu, \phi(x) \rangle$. In this variant, the neuronal drive becomes a linear combination of stored patterns weighted by the gain vector; i.e., the update takes the form $\dot{x} = -x + \Xi p$ (up to timescale factors), and in the fast-gain limit the stationary allocation becomes $p^* \propto \exp(\Xi^\top \phi(x)/T)$, yielding the standard attention readout $x \leftarrow \Xi \text{softmax}(\Xi^\top \phi(x)/T)$ [5,6]. We do not analyze this variant further here; we include it to emphasize that competitive routing on the gain simplex can mechanistically reproduce attentionlike computation under closely related definitions of pattern compatibility.

Numerically, this modulatory degree of freedom improves retrieval accuracy across increasing memory load K and increasing corruption level n (equivalently $\eta = n/N$), out-

performing both the classical Hopfield model [1] and the neuron-astrocyte associative-memory baseline [20] in the tested regimes (Fig. 4). Intuitively, adaptive gain allocation reshapes the effective connectivity (and hence the attractor structure) during recall by amplifying task-relevant patterns and suppressing competitors, in line with recent evidence that stimulus-dependent modulation can improve robustness of Hopfield-type retrieval [28]. This weighted Hebbian decomposition is structurally related to the input-driven Hopfield modulation of Ref. [28], where the weights of stored outer products are determined by the external stimulus. Here, we interpret the weights $p_\mu(t)$ as endogenous dynamical variables: They are constrained to the simplex, evolve through entropy-regularized replicator dynamics, and acquire a softmax fixed point coupled to the current neural state.

The model also suggests qualitative dependencies that could guide future theoretical and computational work. In particular, changes in astrocytic kinetics or background modulatory tone would be expected to alter the sharpness of pattern selection (captured here by T), while retrieval speed and stability should depend on the ratio τ_p/τ_x , setting the competition between routing and neuronal relaxation. More generally, the framework highlights a route by which transient changes in astrocyte-mediated signaling could bias recall without invoking synaptic plasticity. A limitation of the present formulation is that p_μ aggregates multiple biophysical processes and timescales into a single effective variable; identifying which cellular mechanisms and dynamical regimes can realize competitive allocation of this type remains an important direction for future work and connects naturally to multitimescale theories of memory stabilization [29].

An important direction is to convert the proposed competitive routing mechanism into a scalable machine learning architecture, where p acts as a parametrized, context-dependent gating distribution over memory components and T controls gating entropy. This suggests lightweight mixture-of-experts memories with explicit competition and separable routing and state-update timescales (set by τ_p/τ_x). Developing stable, differentiable implementations and benchmarking them on noisy retrieval and continual-learning tasks are natural next steps.

ACKNOWLEDGMENTS

We thank Prof. Luiz Pessoa and Prof. Sergio Gómez for discussions on astrocyte kinetics and associative memory. This work has been supported by Spanish Ministerio de Ciencia, Innovación y Universidades PID2024-158120NB-C21. A.A. also acknowledges the ICREA Academia program of Generalitat de Catalunya.

DATA AVAILABILITY

The code to reproduce the simulations is available in Ref. [30]. There are no publicly available research data or software supporting this manuscript. Requests for further information or data should be sent to the authors.

APPENDIX

For the reader's convenience, we collect here the few definitions that are repeatedly used in the derivations below, so the proofs can be followed without having to refer back to the main text. In particular, the effective synaptic matrix is $W(p) = \frac{K}{N} \Xi \text{diag}(p) \Xi^\top = \frac{K}{N} \sum_{\mu=1}^K p_\mu \xi_\mu \xi_\mu^\top$. For the gain dynamics, $p \in \Delta^{K-1}$ implies $\mathbf{1}^\top \dot{p} = 0$ (tangent-space constraint), and the projector acts on $T_p \Delta^{K-1}$. In the gradient-flow formulation, we use the Shahshahani (Fisher information) metric on Δ^{K-1} , $G(p) = \text{diag}(1/p)$, and the diagonal metric on neuronal preactivations, $H_x = \text{diag}(\phi'(x))$ with $\phi'(x) = \sigma \text{sech}^2(\sigma x)$ for $\phi(x) = \tanh(\sigma x)$.

1. Gradient-flow structure and Lyapunov function

Here, we show that the coupled dynamics in Eqs. (4) and (6) admit a Lyapunov function and can be written as a gradient flow on the product space. We divide the discussion into the astrocyte domain and the neuron domain, where we derive their respective potentials. Note that the coupled dynamics occur in $x, p \in \mathbb{R}^N \times \Delta^{K-1}$ and the metric (needed for the gradient-flow result) is given by a block-diagonal (direct-sum) metric

$$G_{x,p} = H_x \oplus G_p.$$

This metric is composed of the astrocyte and neuron domain metrics. The former is given by the Fisher information metric (equivalently the Shahshahani metric), whose components are

$$g_{\mu\mu'}(p) = \frac{\delta_{\mu\mu'}}{p_\mu},$$

so in matrix form $G(p) = \text{diag}(1/p)$. The latter is given by

$$H_x = \text{diag}(\phi'(x)),$$

where in our case $\phi'(x) = \sigma \text{sech}^2(\sigma x)$.

a. Astrocyte domain

For the astrocyte domain, we first show how the simplex constraints are imposed via the orthogonal projection and then we define the s-replicator potential. With this, we obtain the ODE for the p variable and we discuss its fixed points.

Following Ref. [31], given that we have a differential 0-form (or scalar potential) $\Phi : \mathbb{R}^K \rightarrow \mathbb{R}$ and a metric $G(p)$, we will define an ODE (ordinary differential equation) by computing the gradient of the scalar potential, and then projecting it onto the tangent space of the simplex $\dot{p} \in T_p \Delta^{K-1}$ (i.e., $\mathbf{1}^\top \dot{p} = \sum_{\mu} \dot{p}_\mu = 0$).

In more detail, to compute the gradient, we get the 1-form that lives in the cotangent space $d\Phi \in T_p^* \mathbb{R}^K$; since the gradient lives in the tangent space, using the sharp map musical isomorphism $\#_G : T_p^* \mathbb{R}^K \rightarrow T_p \mathbb{R}^K$, the gradient becomes $z = G^{-1}(p)d\Phi(p)$, where $z \in T_p \mathbb{R}^K$.

Now, we project the ambient/redundant coordinates onto the simplex so that we can express the flow intrinsically in the simplex. To compute the projection onto the tangent space of the simplex $T_p \Delta^{K-1}$ (since we want the evolution law of the ODE to remain on the simplex), we use the orthogonal projector map $\Pi_G : T_p \mathbb{R}^K \rightarrow T_p \Delta^{K-1}$, which takes z to the closest vector in $T_p \Delta^{K-1}$. That is, given a vector z , we solve

for $\Pi_G(z) = \arg \min_{\dot{p} \in T_p \Delta^{K-1}} \|\dot{p} - z\|_G^2$, which returns the vector in $T_p \Delta^{K-1}$ satisfying the minimum distance condition. Written in variational form

$$\mathcal{L}(\dot{p}, \lambda) = \frac{1}{2} \|\dot{p} - z\|_G^2 + \lambda \mathbf{1}^\top \dot{p},$$

its minimum corresponds to the projected vector. Thus, solving for the minimum condition $\nabla_{\dot{p}} \mathcal{L} = 0 = G(\dot{p} - z) + \lambda \mathbf{1} = 0$, we get that

$$\dot{p} = z + \lambda G^{-1} \mathbf{1}. \quad (\text{A1})$$

The condition of a tangent vector to the simplex is $\mathbf{1}^\top \dot{p} = 0$; we obtain

$$\lambda = \frac{\mathbf{1}^\top z}{\mathbf{1}^\top G^{-1} \mathbf{1}}, \quad (\text{A2})$$

where $\mathbf{1}^\top G^{-1} \mathbf{1} = \sum_i p_i = 1$. Plugging λ into Eq. (A1) and using $z = G^{-1} d\Phi$,

$$\dot{p} = [\mathbb{I} - G^{-1} \mathbf{1} \mathbf{1}^\top] G^{-1} d\Phi,$$

and developing the expression, the ODE becomes

$$\dot{p} = (\text{diag}(p) - pp^\top) d\Phi. \quad (\text{A3})$$

Now that we have our gradient-flow ODE, we choose our 0-form to be $\Phi(p) = TD_{\text{KL}}(p||\pi)$, with $\pi = \frac{1}{Z} e^{\frac{1}{T} f}$, such that

$$\Phi(p) = -\langle p, F \rangle + T \ln Z, \quad (\text{A4})$$

where $F = f - T \ln p$ and the differential is given by $d\Phi = T \ln p + T \mathbf{1} - f$ (where neither $\ln Z$ nor f depend on p). Substituting into Eq. (A3), we can write the gradient flow as

$$\dot{p} = -(\text{diag}(p) - pp^\top)(T \ln p - f), \quad (\text{A5})$$

where $T \mathbf{1}$ is proportional to $\mathbf{1}$ and $(\text{diag}(p) - pp^\top) \mathbf{1} = 0$. We refer to this equation as the *s-replicator* (for “soft” replicator), as the entropic term $T \ln p$ prevents winner-take-all dynamics for $T > 0$.

The fixed points of Eq. (A5) satisfy

$$0 = (\text{diag}(p) - pp^\top) F,$$

where $F = f - T \ln p$. One (degenerate) way to satisfy this condition is to make $(\text{diag}(p) - pp^\top) = 0$, where we get two conditions: $p_\mu = p_\mu^2$ and $p_\mu p_\nu = 0$. It can only be satisfied when p is a one-hot vector; thus, the solution lies on the boundary (vertices) of the simplex. Solutions inside the simplex (where $p_\mu > 0 \forall \mu$) satisfy

$$\text{diag}(p) F = pp^\top F.$$

Expressing it elementwise, since $p_\mu > 0$, dividing by p_μ ,

$$0 = F_\mu - \sum_\nu p_\nu F_\nu, \quad \forall \mu,$$

using that $F_\mu = f_\mu - T \ln p_\mu$, and solving for p_μ ,

$$\ln p_\mu = \frac{1}{T} \left[f_\mu - \sum_\nu p_\nu F_\nu \right] \Rightarrow p_\mu = e^{\frac{1}{T} [f_\mu - \sum_\nu p_\nu F_\nu]}. \quad (\text{A6})$$

Imposing $\sum_\mu^K p_\mu = 1$ gives

$$1 = e^{-\frac{1}{T} \sum_\nu p_\nu F_\nu} \sum_\mu^K e^{\frac{1}{T} f_\mu},$$

where taking logarithms, we can identify $\sum_\nu p_\nu F_\nu = T \ln \sum_\mu e^{\frac{1}{T} f_\mu}$ and we can recognize that the partition function $Z = \sum_\mu e^{\frac{1}{T} f_\mu}$. Finally, we can rewrite Eq. (A6) as

$$p_\mu = \frac{1}{Z} e^{\frac{1}{T} f_\mu}, \quad (\text{A7})$$

where in vector form $p^* = \text{softmax}(\frac{1}{T} f^*)$.

b. Neuron domain

Here, we provide a derivation of the classical Hopfield dynamics starting from the energy. We see that the metric $H_x = \text{diag}(\phi'(x))$ emerges as a natural choice. The Hopfield model energy [1] is a scalar function $E : \mathbb{R}^N \rightarrow \mathbb{R}$ expressed as

$$E(x) = -\frac{1}{2} \sum_{ij}^N \phi(x^i) W^{ij} \phi(x^j) + \sum_i^N [x^i \phi(x^i) - L(x^i)], \quad (\text{A8})$$

where $L(x^i) = \int^{x^i} \phi(u) du$ and $\phi(x) = \tanh(kx)$; thus, $L(x^i) = \frac{1}{k} \ln \cosh(\sigma x^i)$. It can be shown that this potential induces a global gradient flow similarly to Ref. [26]. To show this, we compute the differential of the energy $dE = \sum_i \frac{\partial E}{\partial x^i} dx^i$ lying in $dE \in T_x^* \mathbb{R}^N$ and convert it to a gradient flow using the sharp map just like before using the metric $H_x = \text{diag}(\phi')$. Differentiating the first term E_1 , we get

$$\begin{aligned} \partial_{x^k} E_1 &= - \sum_{ij} [\partial_{x^k} \phi(x^i) W^{ij} \phi(x^j) + \phi(x^i) W^{ij} \partial_{x^k} \phi(x^j)] \\ &= - \sum_i \phi(x^i) W^{ki} \phi'(x^k), \end{aligned} \quad (\text{A9})$$

where we have assumed that $W = W^\top$. For the second term E_2 , we get

$$\partial_{x^k} E_2 = x^k \phi'(x^k). \quad (\text{A10})$$

Combining both terms,

$$\partial_{x^k} E = \left[x^k - \sum_i W^{ik} \phi(x^i) \right] \phi'(x^k)$$

or in vector form

$$dE = \text{diag}(\phi') [x - W\phi].$$

Taking the sharp map $\#_H : T_x^* \mathbb{R}^N \rightarrow T_x \mathbb{R}^N$, we get $\dot{x} = -H^{-1} dE$, where the metric necessarily becomes

$$H_x = \text{diag}(\phi'(x)).$$

Note that $1/\phi'(x) > 0$ since in our case $\phi' = \text{sech}^2$. Because H_x is diagonal, it also satisfies $h_x(u, u) = 0$ if $u = 0$ and $h_x(u, v) = h_x(v, u)$ for all x . This is exactly the same as Ref. [26] up to the change of variable $z = \phi(x)$, where z is their dynamical variable; this way, $\dot{z} = \phi'(\phi^{-1}(z))[\phi^{-1}(z) - Wz]$ is exactly $\phi' \dot{x} = \phi'[x - W\phi]$.

c. Joint potential

Finally, we can combine the parts to define the joint potential function $\mathcal{L} : \mathbb{R}^N \times \Delta^K \rightarrow \mathbb{R}$ for the whole system from which we can derive the dynamics as a gradient flow. We define it as

$$\mathcal{L}(x, p) = -KT \ln Z + KTD_{\text{KL}}(p||\pi) + [\langle x, \phi \rangle - L(x)],$$

which combines the two potentials we have described so far. From Eq. (A4), we know that $T D_{\text{KL}}(p||\pi) = -\langle p, F \rangle + T \ln Z$ and we can express it like

$$\mathcal{L}(x, p) = -K\langle p, f \rangle + KT\langle p, \ln p \rangle + [\langle x, \phi \rangle - L(x)],$$

where $f_\mu = \frac{1}{2N}(\sum_i \xi_\mu^i \phi^i)^2$. In this form, the interaction term $\langle p, f \rangle = \frac{1}{2N} \sum_\mu p_\mu \langle \xi_\mu, \phi \rangle^2$ is in direct analogy to $\frac{1}{2} \phi W \phi = \frac{1}{2N} \sum_\mu \langle \xi_\mu, \phi \rangle^2$ of Eq. (A8), where now each term comes multiplied by the modulation p_μ . The second term is the astrocytic regularization and the last one is the neuron activation term. From this potential, we can derive the dynamics of our system by taking the differential living in $d\mathcal{L}(x, p) \in T_{x,p}^*(\mathbb{R}^N \times \Delta^{K-1})$ such that

$$d\mathcal{L}(x, p) = \sum_i^N \frac{\partial \mathcal{L}}{\partial x^i} dx^i + \sum_\mu^K \frac{\partial \mathcal{L}}{\partial p_\mu} dp_\mu. \quad (\text{A11})$$

For the $\partial_p \mathcal{L}(x, p)$ term, it is exactly the same as in Eq. (A5). For the first derivative, we have the interaction and activation terms. The activation term is exactly the same as in Eq. (A10), and for the interaction term we have

$$K \partial_{x^k} \langle p, f \rangle = \frac{K}{N} \sum_\mu^K \sum_i^N p_\mu \xi_\mu^i \xi_\mu^k \phi(x^i) \phi'(x^k),$$

where in vector form $\frac{K}{N} \sum_\mu \xi_\mu^k p_\mu \langle \xi_\mu, \phi(x) \rangle \phi'(x^k) = \text{diag}(\phi') W(p) \phi(x)$, where we define the weight matrix as

$$W(p) = \frac{K}{N} \Xi^\top \text{diag}(p) \Xi.$$

Putting everything together, the two terms of Eq. (A11) become

$$\partial_x \mathcal{L}(x, p) = \text{diag}(\phi') [x - W(p) \phi], \quad (\text{A12})$$

$$\partial_p \mathcal{L}(x, p) = K(T \ln p + T \mathbf{1} - f). \quad (\text{A13})$$

Finally, to get the gradient flow, the metric of the whole space is the direct sum $M(x, p) = H(x) \oplus G(p)$ and the orthogonal projection acts only on the astrocyte subspace; thus, $\Pi = \mathbb{I} \oplus \Pi_G = \mathbb{I} \oplus (\mathbb{I} - G^{-1} \mathbf{1}^\top \mathbf{1} G^{-1})$. The natural gradient descent of $\mathcal{L}(x, p)$ can finally be expressed as

$$\begin{pmatrix} \tau_x \dot{x} \\ \tau_p \dot{p} \end{pmatrix} = - \begin{pmatrix} \mathbb{I} & 0 \\ 0 & \Pi_G \end{pmatrix} \begin{pmatrix} H^{-1} & 0 \\ 0 & G^{-1} \end{pmatrix} \begin{pmatrix} \partial_x \mathcal{L}(x, z) \\ \partial_p \mathcal{L}(x, p) \end{pmatrix}.$$

Thus, the dynamics are as intended:

$$\tau_x \dot{x} = -x + W(p) \phi(x), \quad (\text{A14})$$

$$\tau_p \dot{p} = (\text{diag}(p) - p p^\top)(f - T \ln p), \quad (\text{A15})$$

where we renormalize the time constant $\tau_p/K \rightarrow \tau_p$. This completes the proof that the system is a gradient flow by construction.

2. Dynamical analysis of the system

Now that we have defined the evolution of the system, we can explore both its symmetries and its long-term behavior in different parameter regimes to better understand its dynamics.

3. Symmetry remarks

a. $\mathbb{Z}_2$ invariance of the squared-overlap score

In the main model, the pattern score is a squared overlap, $f_\mu(x) \propto \langle \xi_\mu, \phi(x) \rangle^2$. Consequently, the gain dynamics are invariant under the sign flip $\langle \xi_\mu, \phi(x) \rangle \mapsto -\langle \xi_\mu, \phi(x) \rangle$. Equivalently, defining $m_\mu(x) \equiv \langle \xi_\mu, \phi(x) \rangle$, the gain update depends only on m_μ^2 and cannot distinguish m_μ from $-m_\mu$. More generally, if two patterns satisfy $f_\mu(x) = f_\rho(x)$ at a given state, the gain update has no instantaneous preference between them.

b. Symmetry breaking by neuronal dynamics

Even with the squared-overlap score, degeneracies such as $f_\mu(x) = f_\rho(x)$ are typically lifted by the neuronal evolution. To see this, consider the early-time dynamics with uniform gains $p = \mathbf{1}/K$, for which

$$\tau_x \dot{x}^i = -x^i + \frac{1}{N} \sum_{\mu=1}^K \xi_\mu^i m_\mu, \quad m_\mu \equiv \sum_{j=1}^N \xi_\mu^j \phi(x^j).$$

Differentiating m_μ gives $\dot{m}_\mu = \sum_i \xi_\mu^i \phi'(x^i) \dot{x}^i$; hence,

$$\tau_x \dot{m}_\mu = - \sum_{i=1}^N \xi_\mu^i \phi'(x^i) x^i + \frac{1}{N} \sum_{v=1}^K m_v \sum_{i=1}^N \xi_\mu^i \phi'(x^i) \xi_v^i.$$

Taking the difference between two overlaps yields

$$\tau_x \Delta \dot{m} = - \sum_{i=1}^N \Delta \xi^i \phi'(x^i) x^i + \frac{1}{N} \sum_{v=1}^K m_v \sum_{i=1}^N \Delta \xi^i \phi'(x^i) \xi_v^i, \quad (\text{A16})$$

where $\Delta \dot{m} = \dot{m}_\mu - \dot{m}_\rho$ and $\Delta \xi^i = \xi_\mu^i - \xi_\rho^i$. Except for non-generic trajectories where the right-hand side vanishes identically, $\Delta \dot{m} \neq 0$ and the degeneracy is broken over time. Since $\dot{f}_\mu = (m_\mu/N) \dot{m}_\mu$, this induces a gain reweighting in the astrocytic dynamics. In practice, this mechanism implies that equal-fitness ties are generically transient and are resolved as $x(t)$ evolves away from the initialization.

4. Fixed points of the system

Given the system dynamics of Eq. (A14), fixed-point conditions are

$$x^* = W(p^*) \phi(x^*), \quad (\text{A17})$$

$$p^* = \text{softmax}\left(\frac{f(x^*)}{T}\right), \quad (\text{A18})$$

where the second equation comes from Eq. (A7). Plugging the second equation into the first, it can be rewritten as a single equation on the x domain. In this general regime, there are no analytic solutions, but we can see how the system behaves in different regimes: Either hold the temperature fixed and take asymptotic limits of the time constants or instead fix the time constants and examine the temperature limits.

Case $\tau_p \rightarrow 0$: In this case, we instantly get $p_\mu^* = \text{softmax}(\frac{1}{2NT} \langle \xi_\mu, \phi(x) \rangle^2)$, so the evolution is written as

$$\tau_x \frac{dx}{dt} = -x + W(p) \phi(x).$$

This corresponds to a biased Hopfield network in which every memory is weighted by p_μ^* like $W = \sum_\mu p_\mu^* \xi_\mu \xi_\mu^\top$. Since p_μ^* is highly heterogeneous when $|\langle \xi_\mu, \phi \rangle|$ is high, we are biasing the Hopfield evolution matrix toward those patterns that are more likely to contain the pattern we are searching for.

Case $\tau_p \rightarrow \infty$: If we take $\tau_p \rightarrow \infty$, the astrocytic modulation is frozen: $\dot{p} = 0$; that is, $p(t) = \bar{p}$ is constant. Since the astrocytic influence is initialized by the uniform vector $p = \mathbf{1}/K$, we recover the classical Hopfield dynamics

$$\tau_x \dot{x} = -x + W(\bar{p}) \phi(x), \quad (\text{A19})$$

where $\text{diag}(\bar{p}) = \text{diag}(1/K)$ and thus $\tau_x \dot{x} = -\alpha x + W \phi(x)$.

Case $\tau_x \rightarrow 0$: In this case, the neurons evolve instantaneously to their fixed point $x^* = W(p) \phi(x^*)$. This way, we have

$$\tau_p \dot{p} = (\text{diag}(p) - pp^\top)(f - T \ln p),$$

where $f_\mu = \frac{1}{2NT} \langle \xi_\mu, \phi(x^*) \rangle^2$ is a transcendental function that depends also on p . Since the initial condition for the modulation equation is the uniform $p = \mathbf{1}/K$, the neurons' initial configuration is already at a fixed point of the classical Hopfield dynamic $x^* = W(1/K) \phi(x^*)$. For a sufficiently large number of patterns, the initial fixed point will likely be a spurious minimum. This means, in the very first instants of time, since $x(t_0) = x^*$, the fitness $f(x^*)$ of astrocytic modulation will select those patterns that are the most similar to the $\phi(x^*)$ configuration, which can be arbitrarily far from x_0 and thus even becoming detrimental for retrieval.

Case $\tau_x \rightarrow \infty$: In this limit, the neuron evolution is given by $\dot{x} = 0$, meaning that $x(t)$ is constant it cannot evolve, so we stay forever in the initial condition even if the astrocyte dynamics converge to a fixed point in the p domain.

Case $T \rightarrow \infty$: Keeping now the timescales constant, in this regime, the regularization effect on the s-replicator is so strong that the system stays in the $p = \mathbf{1}/K$ solution. Taking the astrocyte component, we can rearrange for the temperature

to obtain

$$\begin{aligned} \frac{\tau_p}{T} \frac{dp_\mu}{dt} &= \frac{1}{T} p_\mu \left(f_\mu - \sum_v p_v f_v \right) \\ &\quad - p_\mu \left(\ln p_\mu - \sum_v p_v \ln p_v \right). \end{aligned}$$

Taking $T \rightarrow \infty$, we see that $\ln p_\mu = \langle \ln p \rangle$, where we are forced into a maximum entropy distribution. This way we recover the classical Hopfield model.

Case $T \rightarrow 0$: Similarly, rearranging for temperature, if we take $T \rightarrow 0$, the dynamics become

$$\tau_p \frac{dp_\mu}{dt} = p_\mu \left(f_\mu - \sum_v p_v f_v \right),$$

in which case there is no regularization. The fixed points of the system are given by

$$x^* = \frac{1}{\alpha} \Xi \text{diag}(p^*) \Xi^\top \phi(x^*), \quad (\text{A20})$$

$$p^* = \frac{\mathbf{1}_{M(x^*)}}{|M(x^*)|}, \quad (\text{A21})$$

where $M(x)$ is the set of p_μ elements that satisfy $M(x^*) = \arg \max_\mu f_\mu(x^*)$ and $\mathbf{1}_{M(x^*)} \in \{0, 1\}^K$. Note that $M(x^*)$ will generally have a single element unless in the contrived case discussed before in Eq. (A16).

5. Simulations details

To integrate Eq. (6), we use an explicit Euler method. Trajectories are computed for $10/dt$ time steps, with $dt = 0.001$, using smaller steps for small parameter value regimes (i.e., if α is the parameter, then $dt = \alpha \times 0.05$ if $\alpha \leq 0.01$). For statistically significant results, we run 50 simulations with different random pattern matrices Ξ for every different parameter configuration. The observables are computed using the values of the last configuration values $x(t_f)$, $p(t_f)$. To plot the percentile bands, we have smoothed out noise using a 1D Gaussian filter for better visualization.

- [1] J. J. Hopfield, Neural networks and physical systems with emergent collective computational abilities, *Proc. Natl. Acad. Sci. USA* **79**, 2554 (1982).
- [2] D. J. Amit, *Modeling Brain Function: The World of Attractor Neural Networks* (Cambridge University Press, Cambridge, 1989).
- [3] D. Krotov and J. J. Hopfield, Dense associative memory for pattern recognition, in *Advances in Neural Information Processing Systems* (Curran Associates, Inc., Red Hook, NY, 2016), Vol. 29, pp. 1172–1180.
- [4] D. Krotov and J. Hopfield, Large associative memory problem in neurobiology and machine learning, in *9th International Conference on Learning Representations* (2021).
- [5] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, I. Polosukhin, Attention is all you need,

in *Advances in Neural Information Processing Systems* (Curran Associates, Inc., Red Hook, NY, 2017), pp. 5998–6008.

- [6] H. Ramsauer, B. Schäfl, J. Lehner, P. Seidl, M. Widrich, T. Adler, L. Gruber, M. Holzleitner, M. Pavlović, G. K. Sandve, V. Greiff, D. Kreil, M. Kopp, G. Klambauer, J. Brandstetter, and S. Hochreiter, Hopfield networks is all you need, [arXiv:2008.02217](https://arxiv.org/abs/2008.02217).
- [7] A. Araque, R. P. Sanzgiri, V. Parpura, and P. G. Haydon, Astrocyte-induced modulation of synaptic transmission, *Can. J. Physiol. Pharmacol.* **77**, 699 (1999).
- [8] G. Perea, M. Navarrete, and A. Araque, Tripartite synapses: Astrocytes process and control synaptic information, *Trends Neurosci.* **32**, 421 (2009).
- [9] M. Letellier, Y. K. Park, T. E. Chater, P. H. Chipman, S. G. Gautam, T. Oshima-Takago, and Y. Goda, Astrocytes

- regulate heterogeneity of presynaptic strengths in hippocampal networks, *Proc. Natl. Acad. Sci. USA* **113**, E2685 (2016).
- [10] M. De Pittà, N. Brunel, and A. Volterra, Astrocytes: Orchestrating synaptic plasticity? *Neuroscience* **323**, 43 (2016).
- [11] C. Giaume, A. Koulakoff, L. Roux, D. Holcman, and N. Rouach, Astroglial networks: A step further in neuroglial and gliovascular interactions, *Nat. Rev. Neurosci.* **11**, 87 (2010).
- [12] A. Verkhratsky and M. Nedergaard, Physiology of astroglia, *Physiol. Rev.* **98**, 239 (2018).
- [13] S. S. Purushotham and Y. Buskila, Astrocytic modulation of neuronal signalling, *Front. Network Physiol.* **3**, 1205544 (2023).
- [14] S. Makovkin, E. Kozinov, M. Ivanchenko, and S. Gordleeva, Controlling synchronization of gamma oscillations by astrocytic modulation in a model hippocampal neural network, *Sci. Rep.* **12**, 6970 (2022).
- [15] L. Thompson, J. Khuc, M. S. Saccani, N. Zokaei, and M. Cappelletti, Gamma oscillations modulate working memory recall precision, *Exp. Brain Res.* **239**, 2711 (2021).
- [16] K.-I. Dewa, K. Kaseda, A. Kuwahara, H. Kubotera, A. Yamasaki, N. Awata, A. Komori, M. A. Holtz, A. Kasai, H. Skibbe, N. Takata, T. Yokoyama, M. Tsuda, G. Numata, S. Nakamura, E. Takimoto, M. Sakamoto, M. Ito, T. Masuda, and J. Nagai, The astrocytic ensemble acts as a multi-day trace to stabilize memory, *Nature (London)* **648**, 146 (2025).
- [17] S. Zbaranska and S. A. Josselyn, Astrocytes: Emerging stars of engrams, *Cell Res.* **35**, 241 (2025).
- [18] S. A. Josselyn and S. Tonegawa, Memory engrams: Recalling the past and imagining the future, *Science* **367**, eaaw4325 (2020).
- [19] L. Kozachkov, K. V. Kastanenko, and D. Krotov, Building transformers from neurons and astrocytes, *Proc. Natl. Acad. Sci. USA* **120**, e2219150120 (2023).
- [20] L. Kozachkov, J.-J. Slotine, and D. Krotov, Neuron–astrocyte associative memory, *Proc. Natl. Acad. Sci. USA* **122**, e2417788122 (2025).
- [21] P. Coucheney, B. Gaujal, and P. Mertikopoulos, Penalty-regulated dynamics and robust learning procedures in games, *Math. Oper. Res.* **40**, 611 (2015).
- [22] M. A. Di Castro, J. Chuquet, N. Liaudet, K. Bhaukaurally, M. Santello, D. Bouvier, P. Tiret, and A. Volterra, Local Ca^{2+} detection and modulation of synaptic release by astrocytes, *Nat. Neurosci.* **14**, 1276 (2011).
- [23] G. Perea and A. Araque, Astrocytes potentiate transmitter release at single hippocampal synapses, *Science* **317**, 1083 (2007).
- [24] C. Agulhon, T. A. Fiacco, and K. D. McCarthy, Hippocampal short- and long-term plasticity are not modulated by astrocyte Ca^{2+} signaling, *Science* **327**, 1250 (2010).
- [25] E. Shigetomi, S. Patel, and B. S. Khakh, Probing the complexities of astrocyte calcium signaling, *Trends Cell Biol.* **26**, 300 (2016).
- [26] A. Halder, K. F. Caluya, B. Travacca, and S. J. Moura, Hopfield neural network flow: A geometric viewpoint, *IEEE Trans. Neural Networks Learn. Syst.* **31**, 4869 (2020).
- [27] R. A. Jacobs, M. I. Jordan, S. J. Nowlan, and G. E. Hinton, Adaptive mixtures of local experts, *Neural Comput.* **3**, 79 (1991).
- [28] S. Betteti, G. Baggio, F. Bullo, and S. Zampieri, Input-driven dynamics for robust memory retrieval in Hopfield networks, *Sci. Adv.* **11**, eadu6991 (2025).
- [29] M. K. Benna and S. Fusi, Computational principles of synaptic memory consolidation, *Nat. Neurosci.* **19**, 1697 (2016).
- [30] A. Vivet and A. Arenas, Astrocyte-gated associative memory: Code repository, GitHub repository, 2026, <https://github.com/arnauvivet/Astrocyte-gated-associative-memory->, accessed 10 February 2026.
- [31] P. Mertikopoulos and W. H. Sandholm, Riemannian game dynamics, *J. Econ. Theory* **177**, 315 (2018).