

Recognition before Learning: Travelling Waves and Fixed Network Dynamics Support Translation-Invariant Categorization

Hubert Löffler¹

loeffler.hubert@outlook.com

Andreas Bahmer²

Research Article

Keywords: spiking neural network, temporal coding, travelling wave, translation-invariant categorization, hippocampus (CA3), dendritic coincidence detection

Posted Date: October 9th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-11140391/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Recognition before Learning: Travelling Waves and Fixed Network Dynamics Support Translation-Invariant Categorization

Hubert Löffler¹ · Andreas Bahmer²

¹ Independent Researcher, Bregenz, Austria

² RheinMain University of Applied Sciences, Ruesselsheim Campus, 65197 Wiesbaden, Germany

Corresponding author: Hubert Löffler · loeffler.hubert@outlook.com

Keywords: spiking neural network · temporal coding · travelling wave · translation-invariant categorization · hippocampus (CA3) · dendritic coincidence detection

Abstract

Synaptic plasticity is widely regarded as the basis of sensory categorization. Yet network architecture may already provide category-relevant structure before learning. We show that fixed spiking-network dynamics generate population responses that support translation-invariant categorization without synaptic modification. Within the Waves as Space-Time (WST) framework, we built WST-CA3, a spiking network modelled on hippocampal area CA3 (cornu ammonis 3), to study whether spatial and temporal coding together could support learned pattern recognition. In WST-CA3, travelling waves of excitability convert two-dimensional patterns into position-ordered first-spike timing, which mossy-fibre input, recurrent excitation, inhibitory feedback and dendritic delays transform into structured population responses. Unexpectedly, category structure was already present in the untrained network.

With a fixed, centroid-relative read-out of low-order spatiotemporal moments, orientations, blocks, handwritten digits and printed letters were categorized at held-out positions and exemplars: 0.95–1.00 for geometric families (chance 0.17–0.25) and 0.56–0.77 for glyphs (chance 0.10). The read-out can be implemented by spiking circuits, and category cells can self-organize through unsupervised competition. Travelling waves were not strictly required: simultaneous input supported categorization with compact feedforward connectivity, whereas broad connectivity without the wave produced diffuse responses and near-chance performance. Categorization is thus associated with spatially localized population responses. This reveals a further role of travelling waves: they decouple functional from anatomical connectivity, keeping population responses localized despite broad projections. The same architecture also supports temporal pattern separation without dimensional expansion. Fixed architecture and dynamics can therefore provide a translation-invariant, class-separable representation on which subsequent learning can operate.

1. Introduction

How the brain recognizes visual patterns is a central question in neuroscience. Two subprocesses are essential: recognizing a pattern regardless of its position on the sensory surface (translation invariance) and assigning patterns to classes (categorization). In most explanatory models, both depend on learning.

Translation invariance. Recognizing an object regardless of its position is effortless for biological vision but remains a fundamental challenge for neural models. In spatially organized networks, shifting an object changes the neurons and synapses representing it; invariance therefore requires mechanisms that preserve stimulus identity across positions. Two strategies have prevailed. One expands the input into a high-dimensional sparse representation, a role widely attributed to the dentate gyrus (DG) (Marr 1971; O'Reilly and McClelland 1994; Rolls 2016). Sparse expansion does not itself provide invariance but enables trained read-outs to separate stimulus classes. The other develops invariant representations through learning, including convolution and hierarchical pooling (LeCun et al. 1998), predictive coding (Rao and Ballard 1999) and spiking predictive architectures (Lee et al. 2024).

Categorization. Assigning patterns to categories appears to require flexibility, and here too learning is usually taken to be necessary: experience-dependent changes in synaptic weights generate category-relevant representations or decision boundaries, through supervised weight adaptation (LeCun et al. 1998) or through internal representations acquired by predictive and spiking models (Rao and Ballard 1999; Lee et al. 2024). Hippocampal accounts provide sparse, separated representations from which categories can subsequently be learned. Across these approaches, the representation available for subsequent processing is shaped by experience-dependent connectivity. However, object selectivity and even transformation-tolerant object detection can emerge in randomly initialized, untrained hierarchical networks (Baek et al. 2021; Cheon et al. 2022), suggesting that architecture itself can provide substantial structure before learning.

Temporal coding offers one route to such structure. Sensory information is carried not only by firing rates but also by precise spike timing: timing differences can be converted into coincidence by heterogeneous delays (Hopfield 1995), objects may be encoded by the order of first spikes (Thorpe and Gautrais 1997), and spatial visual information is expressed in relative first-spike latencies (Gollisch and Meister 2008). Buonomano and Merzenich (1999) showed that local inhibitory interactions can transform spatial patterns into temporal codes that preserve stimulus identity across positions. In previous studies, we showed that travelling waves of excitability can convert spatial patterns into ordered spike sequences while preserving spatial information (Löffler et al. 2024), and used wave-based transformations between space and time for the memory of spike trains (Löffler and Gupta 2020), temporal pattern separation (Löffler and Gupta 2022) and multiplexing (Löffler and Bahmer 2026).

We previously proposed Waves as Space-Time (WST) as a conceptual framework in which travelling waves transform purely spatial input into temporal output and purely temporal input into spatial output (Löffler and Bahmer 2026). The present study set out to exploit this flexibility for pattern recognition. To this end, we developed WST-CA3, a spiking network built from motifs of hippocampal area CA3 (cornu ammonis 3), in which travelling waves transform two-dimensional patterns into ordered spike sequences. These sequences are processed by a layer of learning neurons (L-neurons) with recurrent excitation, feedback inhibition and biologically plausible dendritic delays. Their recurrent synapses are plastic, allowing categorization to be compared before and after learning.

Unexpectedly, the untrained network already produced category-structured responses: the first-spike clouds of the L-population differed systematically between stimulus classes. This shifted the focus of the study from learning to emergence. Rather than asking how learning enables pattern recognition, we asked whether travelling-wave coding and fixed hippocampal circuitry already provide a representation from which invariant categories can be read out. To remove absolute position, each response is summarized by a small set of centroid-referenced spatiotemporal moments of the spike cloud. We test whether these descriptors support categorization at held-out positions for orientations, blocks, handwritten digits and printed letters; whether they can be decoded by spiking circuits and self-organizing category cells; whether stimulus intensity can be carried within the same code; and whether synaptic plasticity affects the read-out at all.

A second unexpected observation changed our interpretation of the wave itself. When wave-coded input was replaced by simultaneous activation, handwritten digits were still categorized as long as the feedforward projection was compact, whereas with a broad projection the response became diffuse and categorization was no longer distinguishable from chance. The critical factor is therefore not travelling-wave coding per se but a spatially localized population response. Travelling waves provide one biologically plausible mechanism for such localization: they decouple functional from anatomical connectivity, keeping responses localized despite broad projections.

The central claim of this work is therefore not that learning is unnecessary, but that fixed architecture and dynamics already provide a translation-invariant, class-separable representation on which learning can operate.

2. Methods

Methods are organized in two blocks. **Block A (§2.1–2.6)** describes the shared model and measurement machinery used by every experiment: the network, the wave code, the stimuli, the moment features, the category read-out and the evaluation metrics. **Block B (§2.7–2.10)** gives the individual experimental protocols, in the same order as the Results, followed by simulation and reproducibility (§2.11) and the AI-tools statement (§2.12).

Block A — model and measurement machinery.

2.1 Network architecture

Neurons. The model — **WST-CA3** — consists of three interacting neuronal populations on a two-dimensional sheet (here 10×10 units): **input neurons (I)**, **learning neurons (L)** and **inhibitory interneurons (Inh)** (**Fig. 1**). Connectivity follows the canonical hippocampal CA3 motif, comprising feedforward excitation, recurrent excitation and feedback inhibition. I-neurons deliver the stimulus as wave-timed spikes (the dentate mossy-fibre input to CA3). L-neurons are the coincidence-detecting principal cells (dendritic coincidence detection of excitatory postsynaptic potentials (EPSPs) is an experimentally established computation; Stuart and Häusser 2001) whose spike pattern carries the translation-invariant code — they are termed *learning* neurons because their recurrent synapses carry the model's plasticity (§2.8), although a central finding of this study is that this learning can, surprisingly, be omitted for categorization (§3.3, §3.5). Inh neurons provide feedback inhibition. All computation is feedforward unless recurrent connections are explicitly enabled; simulations use a 1 ms integration step.

Relation to the WST framework. WST-CA3 applies the WST principle (Löffler and Bahmer 2026) to a different architecture and task. The original WST model is a one-dimensional, purely feedforward and excitatory

network of four layers — input neurons, oscillation neurons carrying travelling subthreshold membrane-potential oscillations, time-code and space-code neurons, and target neurons — designed so that purely spatial input can be transformed into temporal output and purely temporal input into spatial output, with both forms storable and retrievable. WST-CA3 retains the core mechanism, a travelling wave that converts position into spike timing, but implements it on a two-dimensional sheet with three populations arranged in a CA3-like motif: wave-timed input neurons, coincidence-detecting L-neurons with dendritic branches and pathway-specific delays, and inhibitory interneurons providing feedback inhibition, together with recurrent L → L excitation. It addresses the space → time direction only and asks whether the resulting population response supports translation-invariant categorization.

Connections. The principal excitatory pathway is formed by the projections $I \rightarrow L$ and $L \rightarrow L$. I-neurons provide wave-coded excitation to the L-neurons through a spatially anisotropic Gaussian projection that is narrow along the temporal (x) axis but broad along the spatial (y) axis. The projection is sparse and anisotropic, consistent with the lamellar organization of the mossy-fibre system described in rodents; we do not claim a correspondence between the model axes and the transverse and septotemporal axes of the hippocampus, and no quantitative estimate of mossy-fibre anisotropy is available for the human hippocampus. L-neurons are additionally interconnected by nearly isotropic recurrent excitatory projections, modelling the recurrent collateral network of CA3; during purely feedforward simulations these recurrent connections are gated off (§2.8). Inhibitory feedback is mediated by a separate interneuron population, which combines fast feedforward inhibition from the wave-timed input with feedback inhibition from local principal-cell firing and returns broad, anisotropic inhibition (narrow x, broad y) to the surrounding L-neurons, implementing lateral competition and stabilizing sparse activity (**Table 1**; wiring detail in **Online Resource 1, Sect. S1**).

The high per-synapse gain of the $I \rightarrow L$ (mossy-fibre) pathway relative to the recurrent and entorhinal pathways ($k_I = 3.0$ vs. $k_L = 1.0$, with a weak entorhinal gain k_{EC} that is off in all feedforward simulations; **Online Resource 1, Table S6**) is biologically motivated: the mossy-fibre → CA3 synapse is a conditional "detonator", among the largest and most powerful in the brain, at which even a single active fibre can discharge the postsynaptic cell (Henze et al. 2002; Nicoll and Schmitz 2005). This strong feedforward drive is therefore an architectural rather than a learned property, consistent with the learning-free categorization reported below (§3.3, §3.5). The high gain also lets a sparse, wave-timed input elicit a crisp, short-latency single spike locked to the wavefront, rather than a response assembled by amplitude summation of many weak inputs — which would smear first-spike latency and make firing depend on absolute stimulus intensity. The strong mossy drive therefore cooperates with the wave code rather than conflicting with it; whether the wave is necessary for categorization, or only one way of keeping the response localized, is tested in §2.10. Its proximal placement on the L-neuron dendrite (**Fig. 1b**) likewise follows the anatomy: in the human hippocampus the mossy fibres run through the pyramidal cell layer in the hilus and proximal CA3, turning into the stratum lucidum only near the transition to CA2, which implies a correspondingly proximal, and in part perisomatic and basal-dendritic, termination (Lim et al. 1997). The contrasting biophysics of the weak, modulatory recurrent and perforant-path synapses are detailed in **Online Resource 1, Sect. S1**.

Table 1 Connectivity of the WST-CA3 network

Projection	Function	Geometry	Biological analogue
$I \rightarrow L$	feedforward excitation	anisotropic Gaussian (narrow x, broad y)	DG mossy fibres → CA3
$L \rightarrow L$	recurrent excitation	nearly isotropic Gaussian	CA3 recurrent collaterals
$I \rightarrow Inh$	feedforward drive	one-to-one	feedforward

Projection	Function	Geometry	Biological analogue
			interneurons
$L \rightarrow \text{Inh}$	feedback drive	one-to-one	local interneuron excitation
$\text{Inh} \rightarrow L$	feedback inhibition	anisotropic Gaussian (narrow x, broad y)	lateral / feedback inhibition

Neurons and their connectivity are shown in **Fig. 1**.

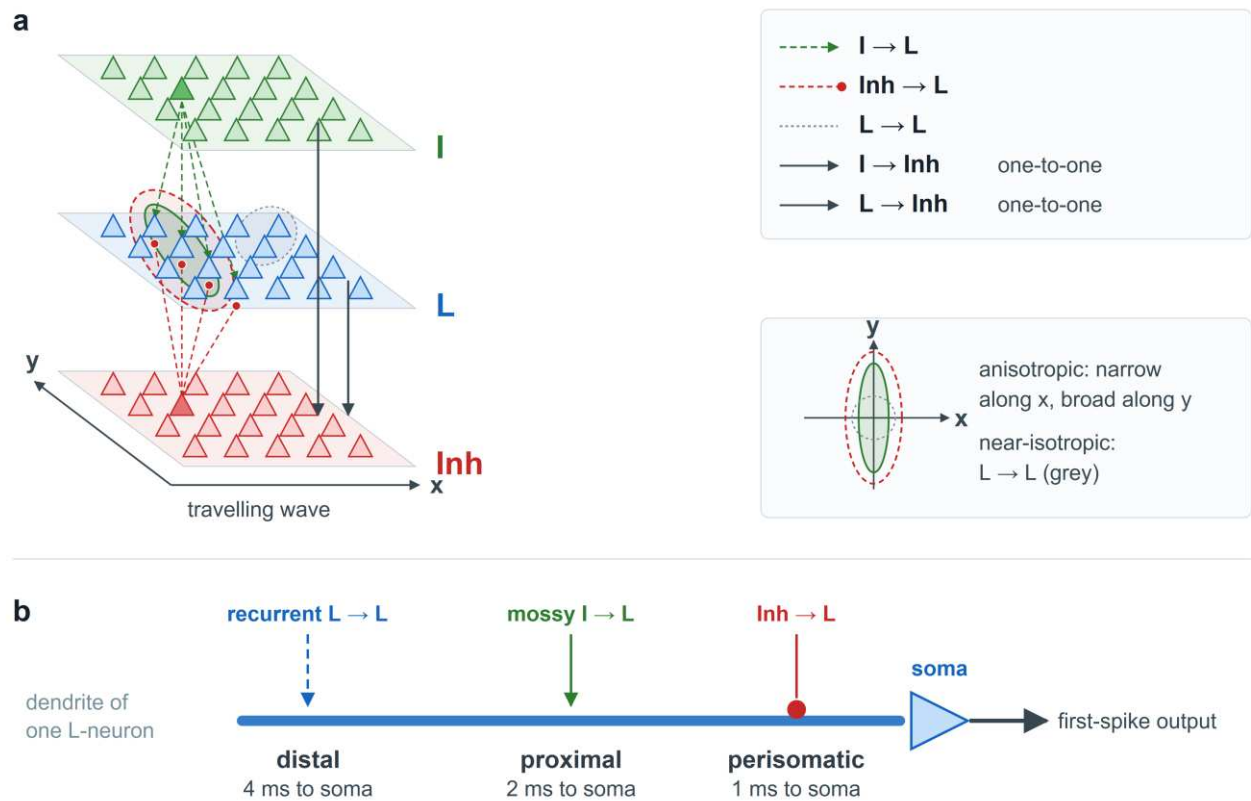


Fig. 1 Architecture of the WST-CA3 network. (a) Three two-dimensional sheets carry the input neurons (I, green), the representation neurons (L, blue; the CA3 pyramidal analogue) and the inhibitory interneurons (Inh, red). Positions are shared across the sheets, so each neuron has a counterpart at the same (x, y) in the other two; the sheets are drawn schematically with fewer units than the simulated 10 x 10. Filled symbols mark one example source neuron. From it, mossy-fibre excitation ($I \rightarrow L$) and feedback inhibition ($\text{Inh} \rightarrow L$) are distributed over anisotropic Gaussian footprints - narrow along the wave axis x and broad along y; the ellipses show one contour of each footprint on the L sheet, enlarged in the inset. Recurrent $L \rightarrow L$ excitation is near-isotropic (grey) and is gated off during feedforward categorization (§2.8). $I \rightarrow \text{Inh}$ and $L \rightarrow \text{Inh}$ are one-to-one, connecting only the neurons at the same position (vertical arrows). The travelling wave advances along x and converts position into first-spike latency (§2.2). In the legend an arrowhead marks excitation and a filled circle inhibition; the footprint shapes are shown in the inset, and all connection types with their biological analogues are listed in Table 1. (b) The same three projections along the dendrite of a single L-neuron: the recurrent input terminates distally (4 ms conduction delay to the soma), the mossy-type input from I proximally (2 ms) and the inhibitory feedback perisomatically (1 ms), so inputs that arrive at the same position reach the soma at different times (§2.1, §2.8). A fourth, most distal region carries the entorhinal perforant path (12 ms) and is used only during pattern completion (§4.4); it is not drawn

Temporal conditions. The wave-encoded input from the I-neurons reaches the dendrites of the L-neurons, where a location-dependent timing rule operates relative to the other inputs arriving there (recurrent $L \rightarrow L$ and inhibitory): the closer an input contacts the soma, the faster it is transmitted. Only the region's delay matters, not its label (**Online Resource 1, Sect. S1**).

The dendritic branch regions of an L-neuron are shown in **Fig. 1b**: inhibitory feedback perisomatically at 1 ms, the mossy-fibre input proximally at 2 ms (plus the dentate relay delay DG_relay) and the recurrent $L \rightarrow L$ input distally at 4 ms. A fourth region, not drawn, the entorhinal perforant path, lies most distally at 12 ms and is engaged only during pattern completion (**Online Resource 1, Sect. S8**); it is inactive in all feedforward simulations reported here.

After an L-neuron spikes, feedback inhibition transiently raises the effective threshold of the other L-neurons (a global membrane downgrade), producing a sparse, competitive response. The full dynamics, time constants, spike/refractory rule and the completion-only recurrent plasticity are given in **Online Resource 1, Sect. S1**.

2.2 Wave coding

The process of wave coding takes place at an earlier stage, present as the wave-based spatial input pattern. The core coding step is a travelling subthreshold-excitability wave that unrolls a static two-dimensional stimulus into a position-ordered sequence of single spikes — expanding the representation **along time** rather than across additional neurons. Each active input site (x, y) emits one spike at

$$t(x, y) = t_0 + x \cdot \Delta t_x + y \cdot \Delta t_y + \xi,$$

with onset $t_0 = beg_Inp$, per-column increment $\Delta t_x = diff_t$, orthogonal increment $\Delta t_y = diff_t_y$, and a small jitter ξ (± 1 ms). With $\Delta t_x = 10$ ms and $\Delta t_y = 0$ the columns fire left-to-right at 10 ms spacing (a gamma-range progression); a non-zero Δt_y tilts the wavefront (**oblique wave**). Biologically the wave is a theta-range subthreshold oscillation that gates one gamma spike per site (**Online Resource 1, Sect. S2**).

Structural desynchronization. Because latency is tied to position, sites sharing an x-column fire **synchronously** while sites spread along x fire **sequentially**: a vertical pattern gives a synchronous volley, a horizontal/diagonal pattern a dispersed one. This position-dependent synchrony transforms geometrically similar patterns into temporally distinct spike sequences (**Fig. 2**). An oblique wave further separates vertically aligned inputs that would otherwise fire synchronously. The wave-ordered spikes then converge onto the L coincidence detectors (brief EPSPs, location-dependent branch delays), whose first-spike pattern — firing positions each with a first-spike time — is the substrate for the moment features (§2.4). Unless stated otherwise: $t_0 = 6$ ms, $\Delta t_x = 10$ ms (gamma ≈ 100 Hz), $\Delta t_y = 0$ or 3 ms, trial duration $nT = 120$ ms (**Online Resource 1, Table S6**). This first-spike latency code follows the same relative-timing principle by which retinal ganglion cells encode a briefly presented image, largely independently of contrast (Gollisch and Meister 2008); its biological grounding and its measured dependence on stimulus orientation are detailed in **Online Resource 1, Sect. S2.1**.

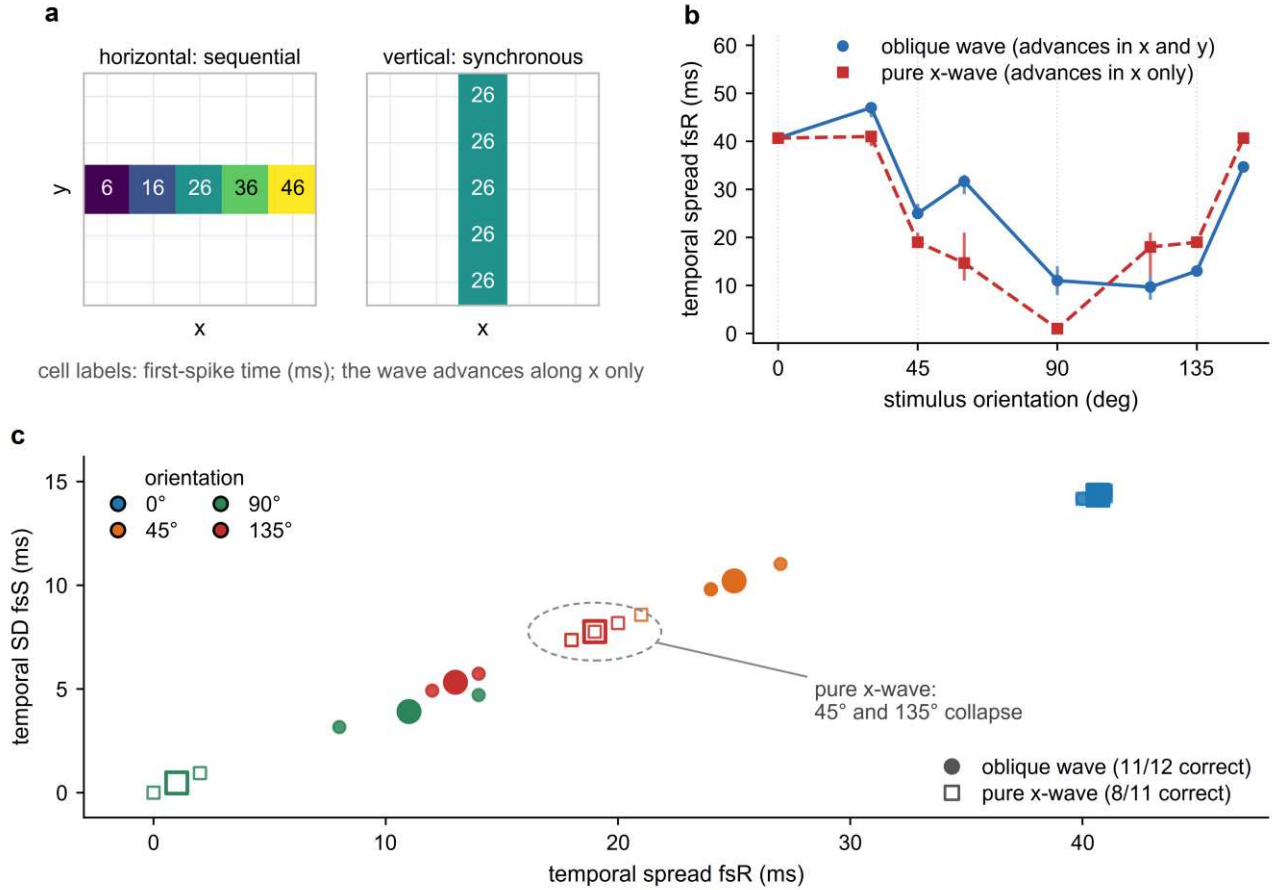


Fig. 2 Wave coding writes spatial configuration into the temporal code. (a) Schematic input sheet. With the wave advancing along x only ($t = t_0 + x \cdot dt_x$, here $t_0 = 6$ ms, $dt_x = 10$ ms), a horizontal line occupies five x-columns and fires sequentially, whereas a vertical line occupies a single column and fires synchronously; cell labels give the first-spike time in ms. (b) Measured in the model, the temporal spread of the L first-spike times (fsR) varies systematically with stimulus orientation, for the oblique wave ($dt_y = 3$ ms) and for a pure x-wave ($dt_y = 0$). Symbols give the mean over the held-out presentations at each orientation and the vertical bars their range; the bar is absent where the values coincide. Three presentations enter each mean, except the pure x-wave at 90 deg, where one presentation produced fewer than two L-spikes and no spread is defined for it. (c) The same held-out presentations in the (fsR, fsS) plane. Under the oblique wave (filled circles) the four canonical orientations occupy separate regions, whereas under the pure x-wave (open squares) 45 and 135 deg collapse into one cluster (dashed ellipse), since they share the same x-projection. A nearest-prototype classifier on (fsR, fsS) alone assigns 11 of 12 held-out presentations correctly under the oblique wave and 8 of 11 under the pure x-wave (chance 1 in 4); all three pure x-wave errors are 45/135 confusions. For a 0 deg stimulus, which has no extent in y, the two wave conditions coincide

2.3 Stimuli and pattern families

The stimuli used as input to the WST-CA3 network were preprocessed using this wave encoding method.

The network is challenged with several families of spatial patterns (**Fig. 3**): two geometric families — orientations and blocks — together with two more complex glyph families, handwritten digits from the MNIST database (Modified National Institute of Standards and Technology) and printed capital letters. Each is treated as an independent classification task within the same predefined moment-based read-out and without shared learning, testing whether one architecture generalizes across pattern types rather than being tuned to a single family. Performance is evaluated on positions and, where applicable, exemplars excluded from prototype construction (§2.6); labels are not used during network processing:

- **Orientations** — a 5-site line through a centre at a given angle. Four canonical angles (0/45/90/135°) form the discrete four-class task; a separate test spans arbitrary angles (0–150° in 30° steps), analysed in §3.3.
- **Blocks** — compact rectangular patches, in two sets that together dissociate the raw and the moment read-outs (analysed in §3.2–§3.3). (i) *Position-differing blocks* — one fixed rectangle (5 × 6 sites) placed at a small set of partially overlapping locations (A_low, B_high, C_right), so the classes differ only in position; they provide the feature-free baseline (§2.7) and the graded-overlap pairs for the pattern-separation analysis (**Online Resource 1, Sect. S5.2**). (ii) *Shape-differing blocks* — four patches differing in aspect ratio and size (wide, tall, small square, large square), each presented across a grid of positions, so the classes differ in shape; this set forms the block translation-invariance family (§2.7).
- **Real-world glyphs** — to test whether categorization generalizes beyond simple geometric patterns, we use the full ten-class MNIST digit set (0–9; LeCun et al. 1998), downsampled and binarized from 28 × 28 to the 10 × 10 input sheet, and the first ten capital letters of the Latin alphabet (A–J), rendered from a standard bold sans-serif font onto the same sheet. Small random rotations ($\pm 10^\circ$) and rescaling (0.85–1.12×) generate shape variants of the letters. Unlike the geometric families, digits and letters combine multiple shape cues and therefore test the fixed read-out under more complex within-class variability (§2.9). Separately from these network simulations, the compression limit of the moment space itself is assessed on the unmodified 28 × 28 images, with the moments computed directly from the pixels and the network bypassed entirely (§3.8).

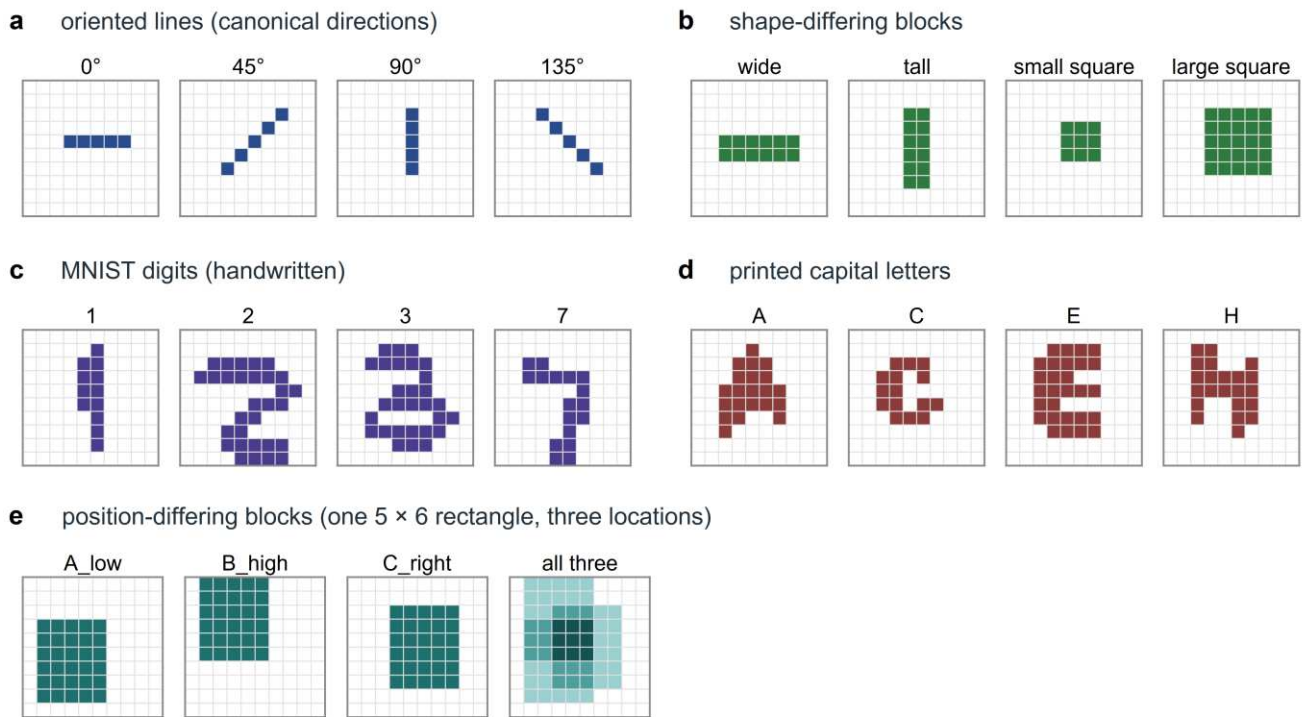


Fig. 3 Example patterns from the five stimulus sets, shown as the network receives them: binary 10 x 10 input sheets, with y increasing upwards. (a) Oriented 5-site lines through a centre, at the four canonical directions; a separate test spans arbitrary angles from 0 to 150 deg in 30 deg steps (§2.3, §3.3). (b) The shape-differing block family: wide (6 x 2), tall (2 x 6), small square (3 x 3) and large square (5 x 5) patches, each presented across a grid of positions. (c) Handwritten MNIST digits, downsampled from 28 x 28 and binarized to the 10 x 10 sheet. (d) Printed capital letters, rendered in a bold sans-serif face and subjected to small random rotations (± 10 deg) and rescaling (0.85–1.12x) before the same downsampling. (e) The position-differing block set used for the feature-free baseline (§2.7, §3.2): a single 5 x 6 rectangle at three partially overlapping locations, so the classes differ in position only. The rightmost sheet superimposes the three, shading each site by how many of them cover it (one, two or all three); this overlap is what makes the set a control for the centroid-relative moment read-out, which cannot see position. Two to four examples are shown per set; the full sets comprise four canonical orientations (a), four block shapes (b), the ten digits 0–9 (c), the ten letters A–J (d) and the three locations (e)

2.4 Spatiotemporal moment features

The L-spike clouds are read out by spatiotemporal moments. In statistics, moments summarize the shape of a distribution by a few numbers: its size (zeroth order), location (first order), spread (second order), asymmetry (third order) and peakedness (fourth order). Applied to the positions and first-spike times of the firing L-neurons, they condense each population response into a small, fixed vector of descriptors — for example, the number of firing neurons, the spatial extent of the spike cloud along x and y, or the spread of its first-spike times. Two descriptors are closely related measures rather than moments in the strict sense: the range of first-spike times (fsR) and the quadrant fractions (rows 11–13).

The moment features are defined here mathematically to provide a transparent read-out of the category information contained in the L-spike clouds. Their computation need not, however, rely on non-neural operations: **Online Resource 1, Sect. S3** shows how the same feature operations can be implemented by biologically plausible spiking circuits.

Table 2 assigns the features to the 13 feature neurons, each with a mathematical quantity, a definition and a label. The default feature set comprises the first **eight** moments (**Table 2**, rows 1–8), carried over from earlier development of the model. Five candidate extension features (**Table 2**, rows 9–13) — the fourth standardized moment (kurtosis) of x_i and of y_i , and the fraction of firing neurons in three centroid-relative quadrants —

were defined so that a feature ablation could test whether higher-order or spatially localized information contributes for any pattern family (§3.4; **Online Resource 1, Sect. S3**).

Table 2 The spatiotemporal moment features: index, symbol, definition, the property each captures, and the functional **family** used for the group ablation (§3.4). Rows 1–8 are the default feature set; rows 9–13 are candidate extensions evaluated in the feature-ablation analysis (**Online Resource 1, Sect. S3.1**). The *tilt* family groups the correlation with the three quadrant fractions, which jointly encode obliqueness.

#	Feature	Definition	Captures	Family
1	n	number of firing L-neurons	mass	count
2, 3	sx, sy	standard deviation of the x-positions and of the y-positions	extent	extent
4, 5	skx, sky	skewness of the x-positions and of the y-positions	direction	skew
6	corr	correlation between the x- and y-positions	orientation (its sign gives the two 45-degree diagonals; together with sx and sy, the arbitrary angle)	tilt
7	fsR	latest first-spike time minus earliest	temporal spread	timing
8	fsS	standard deviation of the first-spike times	synchrony	timing
9, 10	kx, ky	kurtosis of the x-positions and of the y-positions	peakedness vs. spread	kurtosis
11 to 13	q++, q+-, q--	fraction of firing neurons in each centroid-relative quadrant	localized mass	tilt

All spatial descriptors are computed relative to the pattern's own centroid, making them independent of absolute position. Where appropriate, they are additionally normalized by the number of active neurons, preventing response magnitude from trivially scaling with population size. This follows the general principle of moment-based invariant pattern description, in which central (centroid-relative) moments eliminate dependence on absolute position (Hu 1962; Flusser et al. 2009). The first-spike descriptors fsR and fsS extend this classical, purely spatial description into the temporal domain.

Importantly, this mathematical translation invariance of the descriptors is distinct from translation invariance of the network response itself. Centroid referencing removes absolute position from the read-out by construction; it does not ensure that translating an input preserves the category-relevant geometry and timing of the resulting L-spike cloud. Whether the fixed WST-CA3 architecture preserves such structure across stimulus positions is therefore an empirical property of the network, tested below (§3.3).

Exact definitions, including the centroid and the principal-axis orientation, together with the spiking-circuit implementations and worked examples, are given in **Online Resource 1, Sect. S3**.

2.5 Category read-out

The moment features of each L-response are combined into a single feature vector describing the spatial and temporal structure of the L-spike cloud. Classification is therefore based on the complete moment vector, not on individual features. Before classification, each feature is z-standardized using the mean and standard deviation of the reference responses (§2.6), so that features with different numerical ranges contribute on a comparable scale. These values are fixed after the reference set has been presented and do not modify the WST-CA3 network; standardization is thus a fixed gain and offset per feature, realizable in principle by a synaptic weight and a tonic bias.

The primary read-out is a simple **nearest-prototype** classifier. For each category, the moment vectors of a small **labelled** reference set are averaged to form a prototype p_c , i.e. the typical moment vector of that category. A new response is assigned to the category whose prototype is closest to its standardized moment vector z in Euclidean distance ($\hat{c} = \operatorname{argmin}_c \|z - p_c\|^2$). Labels are therefore used only to define the category prototypes; they neither change the WST-CA3 network nor the moment representation.

As a complementary test, we ask whether the same category structure can also be recovered **without labels**. The moment vectors are passed to a competitive downstream layer in which category cells self-organize through winner-take-all competition and local plasticity.

The two read-outs thus serve different purposes. The nearest-prototype rule provides a simple and transparent measure of how much category information is already present in the fixed feature representation. The competitive layer tests whether this structure can also be recovered by an unsupervised neural circuit. The spiking implementations in **Online Resource 1, Sects. S3–S4** are therefore not required for the main result; they provide a proof of principle that the moment features and their categorization can be realized by neural circuitry. Only the per-feature standardization is performed outside the simulated spiking circuits.

2.6 Evaluation procedure and metrics

Procedure. The default network is untrained: no stimulus or category label modifies its weights (§2.1), and learning variants are tested separately (§2.7–2.8). Each **held-out** stimulus is converted into wave-timed input spikes (± 1 ms jitter; N_{VAR} repetitions per position, typically 1–3, see §2.11; §2.2–2.3), propagated through the network in 1 ms steps (§2.11), and its L-response reduced to a moment vector (§2.4).

Reference-position responses are computed only afterwards. They determine the per-feature mean and SD, estimated on the reference set alone and applied unchanged to the held-out responses, and, for the **nearest-prototype** read-out, the mean feature vector of each labelled class (§2.5). The competitive read-out instead self-organizes from the reference responses without labels (**Online Resource 1, Sect. S4**). Labels are therefore used by the primary read-out only to define prototypes; they never modify the WST-CA3 network or its feature representation. The model is called learning-free because no synaptic weight in the network is ever modified — not because this read-out is label-free.

Held-out patterns are never tested at positions used to define their prototypes (and digits and letters also use independent rotation/scale draws). Above-chance performance therefore measures generalization to unseen positions rather than memorization.

The primary metric is **categorization accuracy**, defined for the nearest-prototype read-out as the fraction of held-out responses assigned to the correct class (chance 0.25, or 0.10 for ten-class digits and letters). For the competitive layer, we report the number of **distinct** dominant category cells (ideally one per class) and their **purity** — the fraction of held-out responses assigned to their class's dominant cell.

Pattern separation and coding cost are defined in **Online Resource 1, Sect. S5.2**. Unless stated otherwise, results use the pinned seed (§2.11). Translation-invariance results (§3.3) additionally report 95 % Wilson confidence intervals, variation across three seeds, and label-permutation baselines.

2.7 Translation-invariance protocol

Having been surprised by categorization performance without any learning, we set out to determine whether it is in fact due to the fixed wiring. Three separate questions are involved, and the protocol addresses them one at a time: whether the raw population response already carries category structure, whether that structure survives translation to unseen positions, and whether the moment read-out is genuinely blind to position.

A full description of these three tests — the raw-response separation measures, the leave-one-position-out protocol with its admissibility rule and control splits, and the position-decoding test — is given in **Online Resource 1, Sect. S10**, together with their results (summarized in §3.3). These tests formed the basis for the further investigation of categorization performance in WST-CA3.

2.8 Can learning improve categorization?

Here we return to the original aim of the study: whether synaptic learning within WST-CA3 can improve categorization. The default network is untrained, and its only plastic pathway is the recurrent $L \rightarrow L$ projection (§2.1); learning in the downstream competitive layer is a separate matter (§2.5; **Online Resource 1, Sect. S4**). Because categorization is read out from the L-response, recurrent plasticity can improve it only if it alters that response. We therefore test this reachability directly. **The recurrence gate.** Recurrent input is admitted only while $(T - T_0) \leq \text{RECUR_T_MAX}$ (the width of the retrieval window, in 1 ms simulation steps), where T_0 is the step immediately preceding the first input spike of the trial,

$$T_0 = \min(x) \cdot \Delta t_x + t_0 - 1.$$

In the canonical configuration $\text{RECUR_T_MAX} = 0$. The first L-spikes occur about 4 ms after T_0 , and recurrent transmission adds a further 4 ms of dendritic delay, so recurrent spikes cannot influence the initial L-response under the canonical gate. Empirically, any setting up to 8 blocks recurrent transmission completely, and transmission first appears at about 10.

Reachability of recurrent learning. We test directly whether plasticity in this pathway can alter the L-response, rather than inferring its effect from learning-rate sweeps. The recurrent synapses are trained on reference stimuli for three and for ten epochs under the balanced rule (homosynaptic and heterosynaptic potentiation and depression all at 0.5: $\text{LTP_L} = \text{LTD_L} = s_{\text{LTP_L}} = s_{\text{LTD_L}} = 0.5$). Learning is then switched off and the trained and the untrained network are presented with the same held-out stimuli; their L-spike matrices are compared spike by spike, so that any effect of recurrent learning on the response is detected without relying on a downstream summary measure.

Two further measurements establish the strength and the accessibility of the recurrent pathway: how far training moves the mean recurrent weight, and how much uniform amplification of the recurrent branch is

required before the L-response changes at all — scanned from the canonical value to fiftyfold, both at the canonical gate and at an opened one. If the trained and untrained L-responses are bit-identical, their moment vectors and categorization outputs must be identical too, and a separate accuracy comparison would add no information.

2.9 Generalization to real-world digits and letters

MNIST digits 0–9 and capital letters A–J (§2.3) test whether the same fixed architecture and translation-invariant moment read-out generalize from simple geometric patterns to complex glyphs. The glyphs are represented by the extended 13-feature set (§2.4) and evaluated on held-out exemplars — for the letters, on rotation and scale draws never used to build the class descriptors (§2.6). Chance is 0.10.

Because the glyphs are substantially denser than the geometric stimuli, they run under a single **stated deviation** (inhibitory gain $k_{Inh} = 2$, $diff_t = 5$ ms, $nT = 70$ ms; **Online Resource 1, Table S6**), all other parameters as pinned. On the 10×10 sheet a glyph fills almost the whole field, so this protocol varies exemplar and shape but not position; translation for the glyph families is therefore tested separately, on an enlarged sheet with further stated deviations (§3.5). Performance is reported as overall held-out accuracy together with the confusion matrix, so that the structure of the remaining errors is visible and not only the overall rate.

2.10 Is travelling-wave coding necessary? Wave versus footprint

Category structure is present in the untrained network, but which of its properties produces it has not yet been isolated. Wave-coded input is the defining feature of WST-CA3, yet other structural properties could be equally decisive — in particular the width of the feedforward and inhibitory projections, which determines how far the response to a local input spreads across the L-sheet.

The preceding experiments use travelling-wave input throughout. To determine whether the wave itself is necessary for categorization, we separate its effect from that of feedforward spatial localization. Both mechanisms can restrict the L-population response: a compact $I \rightarrow L$ projection does so anatomically, whereas a travelling wave can do so temporally even when the anatomical projection is broad.

The temporal mechanism follows from the interaction between wave timing and the brief mossy coincidence window. Because position along the wave axis is converted into input latency (§2.2), afferents distributed across a broad $I \rightarrow L$ footprint do not arrive simultaneously, and only inputs originating within a sufficiently narrow temporal interval summate effectively. The ratio of the per-column wave delay to the mossy coincidence timescale, $\Delta t_x / \tau$, therefore gives an approximate measure of the temporal sharpening along the wave axis: about eightfold under the canonical $\Delta t_x = 10$ ms, and about fourfold in the dense-glyph configuration used for the test below ($\Delta t_x = 5$ ms; $\tau \approx 1.2$ ms; **Online Resource 1, Sect. S1**). It measures temporal sharpening, not a literal anatomical reduction factor.

We test the two mechanisms independently, using the real-world digit configuration of §2.9. Wave coding is either retained or replaced by simultaneous input ($diff_t = 0$), while the width of the feedforward projections is varied from compact to broad ($with_I/L/Inh$; §2.1), with the connectivity rebuilt for each footprint. For each wave $\times$ footprint condition we measure, in the same run, held-out categorization accuracy (§2.6), the number of firing L-neurons, and — per presentation — the spatial extent of the L-response along the wave axis together with its temporal spread. This factorial manipulation tests whether successful categorization specifically requires travelling-wave coding or, more generally, a sufficiently localized L-population response.

2.11 Simulation and reproducibility

The model is implemented in Python (NumPy) at 1-ms resolution on a 10×10 neuronal sheet, with a canonical trial duration of $nT = 120$ ms. Experiment-specific deviations from the canonical configuration — including sheet size and trial duration — are stated in the final column of **Online Resource 1, Table S6**. A batch/importable model, used for all quantitative experiments, and an interactive spreadsheet "cockpit" share the same neuronal and synaptic dynamics; a frozen copy of the cockpit defines that canonical configuration. To eliminate any residual dependence on the live state of the interactive implementation, each consolidated result is generated by a script that explicitly pins the complete parameter set.

Presentations use seeded timing jitter (`numpy.default_rng`; seed 7 unless otherwise stated). Each class is shown at each admissible position in N_VAR jittered variants, N_VAR being experiment-dependent (typically 1–3; four in the principal translation-invariance analysis of §3.3; **Online Resource 1, Table S6**). Given the same seed and stimulus specification, simulations are deterministic. Each reported figure and quantitative result can be regenerated from the script listed for it in **Online Resource 1, Sect. S6**. Complete implementation details, integration procedures, evaluation protocols and runtime characteristics are provided in **Online Resource 1, Sect. S5**.

2.12 Use of AI tools

During the preparation of this work the authors used Claude (Anthropic) and ChatGPT (OpenAI) to assist with drafting and editing parts of the manuscript, implementing and running supplementary analyses and control experiments, computing the reported interval estimates and significance tests, optimizing research code, creating schematic figures, and checking bibliographic references. All AI-assisted analyses, code modifications, figures, and manuscript content were subsequently reviewed by the authors. The authors take full responsibility for the methods, results, interpretations, and final content of the publication.

3. Results

All results in §3.2–§3.6 use the fixed baseline configuration `config_4_01` (Online Resource 1, Table S6), except for the explicitly stated dense-glyph condition used for MNIST digits and printed letters (§3.5), for which three input-processing parameters are adjusted ($k_inh = 2$, $diff_t = 5$ ms, $nT = 70$ ms; §2.9) while the network architecture remains unchanged. Feature sets are defined independently in §2.4: the geometric tests use the default eight-feature set, whereas the complex glyphs use the extended 13-feature set evaluated by ablation in §3.4. Within each experiment, all reported comparisons use the same stated configuration. We first show how travelling-wave coding transforms spatial input geometry into a separable temporal code (§3.1). We then report the central, unexpected observation: even before feature extraction or learning, the fixed, untrained network transforms overlapping inputs into distinctly structured population responses (§3.2). To determine whether this structure is preserved independently of stimulus position, we apply the translation-invariant moment read-out and test categorization across held-out positions (§3.3), identify which features contribute to each pattern family (§3.4), and examine whether the same fixed pipeline generalizes to handwritten digits and printed letters (§3.5). We then ask whether synaptic learning can improve this categorization, and find that the model's only plastic pathway cannot reach the read-out at all (§3.6). We further show that the read-out can be implemented by biologically plausible circuits (§3.7) and summarize additional operations supported by the model — temporal separation and intensity coding — together with its coding cost (§3.8). Finally, we revisit whether the travelling wave is itself necessary, finding that response localization, rather

than wave coding per se, predicts categorization (§3.9). Pattern completion, a second operation of the recurrent pathway, is considered in the Discussion.

3.1 Wave coding produces a separable temporal code

The input to the WST-CA3 network is a spatiotemporal pattern generated by wave coding. The travelling wave assigns each active site a first-spike latency determined by its position ($t = t_0 + x \cdot \Delta t_x + y \cdot \Delta t_y$; §2.2). Sites sharing an x-column fire synchronously, whereas sites distributed along x fire sequentially, so that a vertical pattern produces a synchronous volley and a horizontal or diagonal pattern a temporally dispersed one. Measured directly in the model, this temporal spread — the range and standard deviation of the L first-spike times (fsR , fsS) — varies systematically with stimulus orientation and places the four canonical orientations at distinct positions in the timing plane, so orientation is written into the temporal moments as a complementary channel to the spatial tilt (§3.4). Different spatial configurations thereby acquire distinct temporal signatures that downstream coincidence detectors can read. The oblique component ($\Delta t_y \neq 0$) additionally desynchronizes vertically aligned inputs that a pure x-wave would leave simultaneous — and likewise separates the 45°/135° pair, which a pure x-wave leaves temporally near-identical: all three errors of a nearest-prototype classifier on the timing pair are then 45°/135° confusions (8 of 11 held-out presentations correct, versus 11 of 12 under the oblique wave; chance 1 in 4) — which becomes particularly important for the near-vertical orientations examined in §3.3. These timing effects are illustrated in **Fig. 2**.

3.2 The untrained network already separates classes

Before introducing moment-based feature extraction or synaptic learning, we asked whether the raw L-population response already contains class-separable structure. The three block classes used here (A_{low} , B_{high} and C_{right} ; §2.3) are copies of the same rectangle placed at three partially overlapping positions. They therefore differ in location but not in shape, making this an intentionally position-confounded test of whether the fixed network transforms similar, overlapping inputs into distinguishable population responses. Each class was presented as noisy variants at its own position. Separation was computed on four held-out variants per class (12 responses in total), each represented in two raw forms — the spatial firing pattern (which L-neurons fired) and the corresponding first-spike latencies — as Δ = mean within-class similarity – mean across-class similarity over all pairs of responses, together with the silhouette coefficient (Rousseeuw 1987). The silhouette asks, for each response, whether it lies closer to the other members of its own class than to the members of the nearest other class. With a the mean distance of a response to its own class and b its mean distance to the nearest other class (distance = 1 – similarity), its silhouette is $s = (b - a) / \max(a, b)$; s is averaged over all responses. Values near 1 indicate compact, well-separated classes, values near 0 overlapping classes, and negative values responses that lie closer to another class than to their own. Unlike Δ , which averages over all pairs of classes, the silhouette compares each response with its nearest competing class and is therefore the stricter measure when two classes lie close together (details in **Online Resource 1, Sect. S10**).

The fixed, untrained network separates the three classes in both raw codes, modestly but systematically (**Fig. 4**): the spatial code (which L-neurons fired) reaches $\Delta = 0.25$ (silhouette 0.27) and the latency code (first-spike times) $\Delta = 0.11$ (silhouette 0.10). The separation recurs across three seeds and exceeds a label-permutation baseline in every case. Silhouette values of this size indicate overlapping rather than compact classes, particularly in the latency code. Thus, without feature extraction or synaptic learning, same-class variants evoke more similar population responses than variants from different classes, more strongly in the spatial than in the temporal code. This raw separation does not require travelling-wave timing: with simultaneous

input it persists, because the classes differ in position and the identity of the firing L-neurons already carries that difference (**Online Resource 1, Sect. S10**).

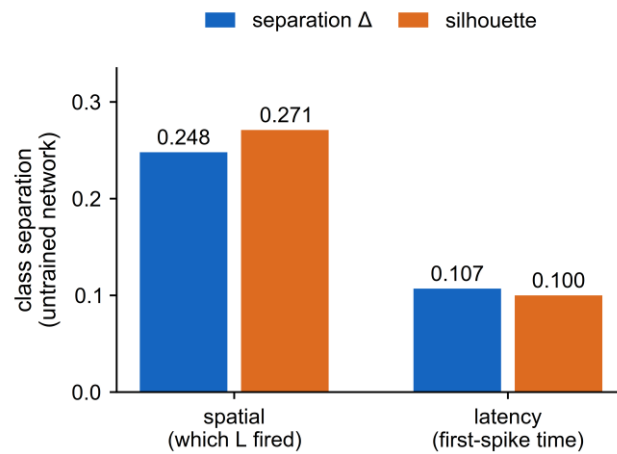


Fig. 4 Raw class separation in the untrained network. Separation Delta (mean within-class minus mean across-class similarity) and the silhouette coefficient, for the raw spatial firing pattern (which L-neurons fired) and for the first-spike-latency code, on the three overlapping position-defined block classes (A_low, B_high, C_right; Fig. 3e). Evaluation uses four held-out noisy variants per class, 12 responses and 66 pairs in total, at the same class-defining positions. Positive values indicate greater similarity within than between classes. No feature extraction and no synaptic learning are involved. Values for seed 7; statistics across seeds 7, 11 and 13 and the control without the travelling wave are given in Online Resource 1, Sect. S10. Because these classes differ in position rather than in shape, this separation does not yet demonstrate translation-invariant categorization; §3.3 and Fig. 5 test that separately

This separation does not yet demonstrate translation-invariant categorization, because class identity is confounded with stimulus position and the raw spatial response can directly retain positional information. It does, however, show that the fixed network transforms overlapping noisy inputs into systematically structured population responses before learning. Section 3.3 therefore asks the stricter question: does category structure remain when absolute position is removed from the read-out?

Raw Δ is not used to assess the effect of learning. It is an overlap measure and is therefore sensitive to the density of the response; a change in Δ alone does not establish improved categorization. Section 3.6 instead tests the prior question of whether recurrent plasticity — the network's only plastic pathway (§2.8) — changes the L-spike response at all.

3.3 Translation-invariant categorization across pattern families

Centroid-relative spatiotemporal features remove absolute response position from the read-out, but translation invariance of the complete network response remains an empirical question because wave timing and anisotropic connectivity can alter response geometry across positions. We therefore tested categorization by leave-one-position-out evaluation (§2.7), with every admissible position held out in turn.

Held-out accuracy was .997 for canonical orientations, .954 for arbitrary orientations, and .953 for shape-defined blocks (chance .25, .17, and .25, respectively; **Fig. 5**). Performance was essentially unchanged when variation was introduced by position rather than stimulus variant, whereas absolute position could not be decoded above chance from the same representation (**Online Resource 1, Sect. S5.1**); the full statistics — sample sizes, confidence intervals, the three-seed spread and the permutation baselines — are given in **Online Resource 1, Sect. S10**. Thus, translation invariance was not merely a mathematical consequence of

centroid-relative descriptors but remained present across the complete input→network→read-out transformation.

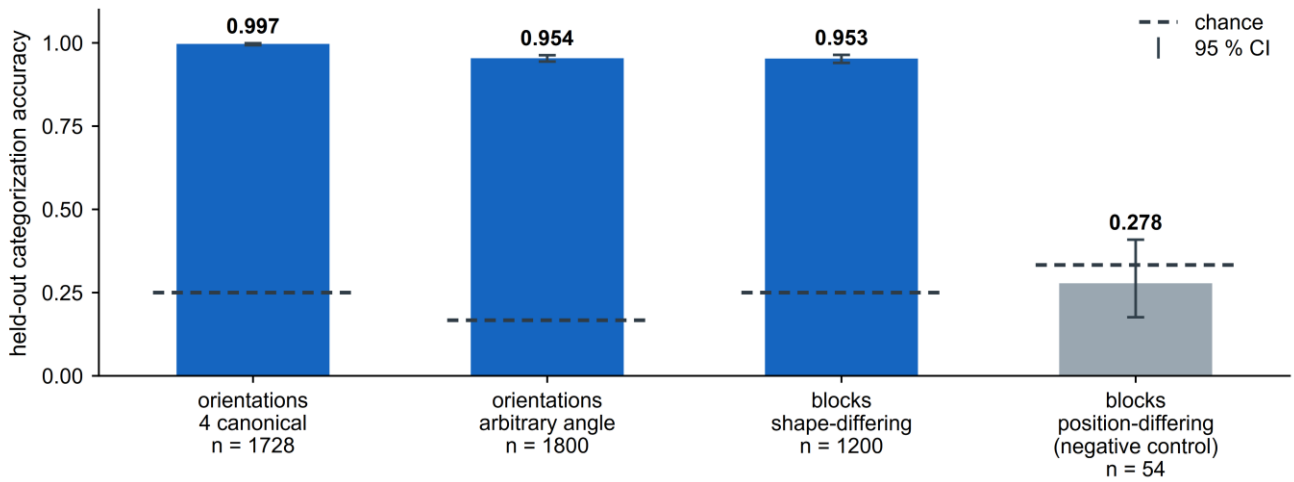


Fig. 5 Translation-invariant categorization by the moment read-out at held-out positions, for four pattern families. Bars give held-out accuracy under the leave-one-position-out protocol of §2.7, whiskers the 95 % confidence interval, and the dashed segment the chance level of that family (0.25, 0.17, 0.25 and 0.33). Sample sizes are given under each family. The position-differing blocks (grey) are the negative control: three copies of one rectangle that differ only in location, so a centroid-relative read-out has no shape distinction to find. Their interval spans the chance level, whereas the three shape-defined families lie far above it. Full statistics, including the three-seed spread and the permutation baselines, are in Online Resource 1, Sect. S10

3.4 Different pattern families rely on different moment features

Identifying which features carry category information makes the representation mechanistically interpretable. To determine the contribution of individual feature families to categorization, each family was removed in turn from the complete feature representation while all other features and evaluation procedures were left unchanged. The resulting decrease in held-out accuracy quantifies the information uniquely or non-redundantly contributed by that feature family. We removed one functional feature family at a time (**Table 2**) and measured the resulting decrease in held-out classification accuracy (**Fig. 6**). Families rather than individual features were ablated because features within a family can carry partly redundant information; removing a single feature may therefore underestimate the contribution of the underlying property. The ablation is run on the extended 13-feature set for every pattern family, so that the same six feature families can be compared throughout; for the geometric families it uses the leave-one-position-out split of §2.7. Its baselines therefore lie marginally above the values in §3.3, which are measured under the same split but on the default eight-feature set (arbitrary angles 0.962 versus 0.954; canonical orientations and shape-differing blocks differ by 0.003 or less).

orient. canonical (base 1.00)	0.00	0.00	0.00	0.23	0.00	0.00
orient. arbitrary (base 0.96)	0.00	0.01	0.00	0.31	0.01	0.00
blocks position (at chance)	0.05	-0.05	0.05	0.08	-0.03	0.02
blocks shape (base 0.95)	-0.01	0.09	-0.01	-0.01	0.00	0.00
letters A–J (base 0.83)	0.07	0.01	0.11	-0.01	-0.03	0.11
MNIST 0–9 (base 0.57)	0.02	0.01	0.05	0.02	-0.02	0.06
	count	extent	skew	tilt	timing	kurtosis
moment feature family removed						

accuracy drop when removed

- (< 0.03, noise)
- weak (0.03–0.10)
- strong (≥ 0.10)
- at chance: not informative

Fig. 6 Which moment family carries which pattern (group ablation). Each cell gives the drop in held-out accuracy when that feature family is removed from the complete 13-feature set; the colour marks the band (< 0.03 noise, 0.03–0.10 weak, ≥ 0.10 strong). Tilt carries orientation, both canonical and arbitrary; extent carries the shape-differing blocks, though less strongly; letters draw on skew and kurtosis; MNIST digits show no single strong carrier. The position-differing row is hatched: those classes are already at chance (§3.3), so removing a feature family from them is uninformative. Rows are not directly comparable: the geometric rows use the leave-one-position-out split of §2.7 and resolve one item to about 0.001, whereas the glyph rows use the glyph protocol of §2.9 and resolve one item to about 0.01. Baselines are those of the 13-feature set and therefore lie marginally above the eight-feature values of §3.3 (arbitrary angles 0.962 versus 0.954). Exact values are in Online Resource 1, Table S5

Different pattern families are thus separated by different, interpretable feature subspaces: no single feature family accounts for all tasks, but the small fixed set spans the distinctions required across them. The per-family analysis — which moment family carries which pattern — together with further ablation analyses, feature extensions and controls, is reported in **Online Resource 1, Sect. S3.1**.

3.5 Generalization to handwritten digits and printed letters

To test whether the results obtained with geometric stimuli extend to more complex patterns, we applied the same fixed architecture and moment read-out to handwritten MNIST digits 0–9 (binarized to the 10×10 input sheet) and printed capital letters A–J (rendered with small rotation and scale jitter). Both are ten-class problems (chance = 0.10). Because these glyphs are substantially denser than the geometric stimuli, they were evaluated under the predefined dense-input configuration ($k_{\text{Inh}} = 2$, $\text{diff}_t = 5$ ms, $nT = 70$ ms; §2.9), with all remaining parameters fixed. For these more complex glyphs, the read-out uses the extended 13-feature set defined in §2.4; the contribution of the extension features is assessed independently by ablation in §3.4.

The untrained network categorizes both stimulus sets well above chance: accuracy reaches 0.59 for MNIST digits and 0.83 for printed letters (**Fig. 7**). The confusion matrices show structured rather than uniform error patterns consistent with visual similarity. Among the digits, classes 0, 1, and 6 each reach 0.90 accuracy, whereas errors are concentrated in more frequently confused classes, including $2 \rightarrow 0$ and $9 \rightarrow 4/7$. Among the letters, most classes reach at least 0.90 accuracy, with errors concentrated particularly on the visually similar rounded forms B and D. These results show that the category structure observed for the geometric stimuli extends to more complex glyphs and is not restricted to the synthetic pattern families used in §3.3.

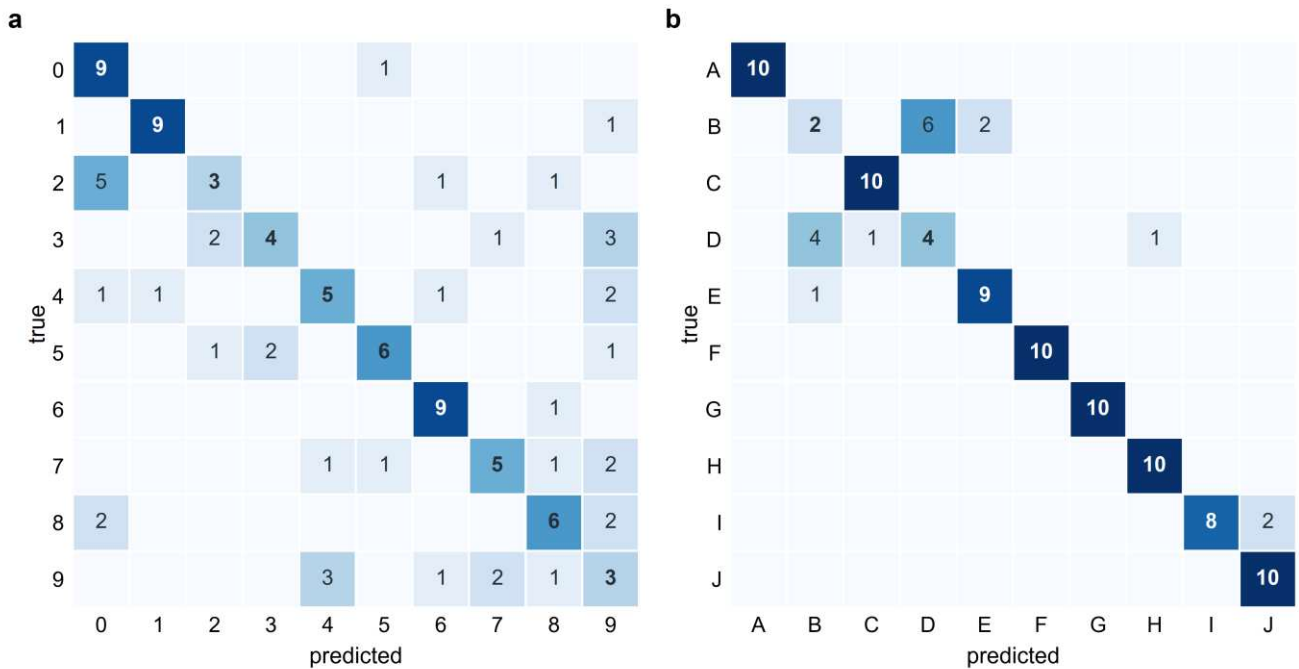


Fig. 7 Generalization to complex glyphs. Confusion matrices for handwritten MNIST digits 0-9 (a) and printed capital letters A-J (b), using the 13-feature read-out under the dense-glyph configuration of §2.9. Rows are true classes, columns predicted classes; empty cells are zero and the diagonal is set in bold. Each class contains 10 held-out items, so the diagonals give 59 of 100 (0.59) and 83 of 100 (0.83) correct classifications against a chance level of 0.10. The errors are structured rather than uniform: among the digits they concentrate on 2 to 0 and 9 to 4 or 7, among the letters on the rounded pair B and D

Translation invariance of all pattern families — the three tests of §2.7, the coverage controls and the leave-one-offset-out test of the glyphs on an enlarged sheet — is examined in **Online Resource 1, Sect. S10**.

3.6 Recurrent plasticity cannot reach the categorization read-out

The only plastic pathway in the model is the recurrent $L \rightarrow L$ projection (§2.1). Whether learning there can alter the categorization read-out is decided by the architecture rather than by the learning rule, because the gating window of the canonical configuration closes before any recurrent spike can arrive: the first L-spikes occur about 4 ms after the first input spike, and the recurrent branch adds a further 4 ms of dendritic delay (§2.8).

Learning nevertheless takes place, and not marginally. Three epochs of the balanced rule lower the mean recurrent weight by 14.9 % and ten epochs by 22.9 %, while the feedforward mossy weights remain untouched. Scoring the trained network on the eight held-out presentations — the four block shapes at two centres never used for training — then returns a response identical to the untrained one **spike for spike** — the same L-neurons firing at the same first-spike times — after three epochs and after ten. The read-out receives exactly the same input before and after training, so no accuracy comparison can distinguish the two.

The gate, not the size of the learned change, is what blocks the pathway. Amplifying the recurrent weights uniformly leaves the response unchanged up to fiftyfold (68 L-spikes across eight held-out presentations at every gain from $\times 1$ to $\times 50$). Opening the window to $\text{RECUR_T_MAX} = 10$ makes the same untrained pathway effective immediately: the response grows to 78 spikes, and further amplification changes it monotonically (118 spikes at $\times 1.5$, 349 at $\times 50$).

Recurrent plasticity therefore has no influence on any categorization result reported here, and §3.3–§3.5 are independent of learning by construction rather than by measurement. This is a statement about the configuration in which categorization is scored, not about plasticity in general: whether learning contributes once the recurrent window is open — the regime in which the same pathway supports pattern completion (§4.1; **Online Resource 1, Sect. S8**) — remains untested.

3.7 A spiking-circuit implementation reproduces the analytical read-out

To test whether the moment read-out requires explicit mathematical computation or can be obtained by biologically plausible means, we replaced the closed-form features with spiking circuits that approximate the same centroid-relative moments (§2.4) through simple neural operations — pooling, squared and opponent pools, product interactions, coincidence detection and temporal pooling (**Online Resource 1, Sects. S3–S4**). Categorization was essentially preserved. For four orientations at held-out positions, the circuit-derived features classified 71 of 72 presentations correctly (0.99; 95 % Wilson CI [0.93, 1.00]; seeds 7, 11 and 13, per seed 0.96–1.00), compared with 72 of 72 (1.00) for the analytical features (chance 0.25). Individual circuit motifs matched their analytical counterparts: the skew-opponent motif reached 71/72 (0.99) with both feature types in a four-class task, and the product/diagonal and coincidence/synchrony motifs 36/36 each (1.00; CI [0.90, 1.00]) in two-class tasks (chance 0.50).

3.8 Additional computations and coding economy

Beyond spatial categorization, wave-based encoding supports further computations, examined here at different depths.

Temporal pattern separation. For spatially overlapping 3×3 blocks (16 patterns; 90 pairs sharing 1–6 of 9 sites, mean input overlap $OI = 0.20$; seeds 7/11/13), wave-coded input decorrelated the L-responses in the temporal domain (separation power, $SP = +43 \pm 5 \%$, mean $\pm$ SD across seeds), although their spatial population overlap increased ($SP = -70 \pm 5 \%$; $SP < 0$ when the output overlaps more than the input, **Online Resource 1, Sect. S5.2**). Without the wave, the temporal channel reversed sign ($SP = -80 \pm 6 \%$). Separation is therefore carried by first-spike timing rather than by expansion into a larger neuronal population. This extends the one-dimensional temporal separation reported previously (Löffler and Gupta 2022) to two-dimensional spatial patterns and complements temporal pattern separation observed experimentally in hippocampal neurons (Madar et al. 2019).

Intensity coding. Stimulus intensity can be added to the same representation by scaling the strength of the synaptic event while leaving input spike timing unchanged. All eight combinations of four intensity levels at two held-out positions were classified correctly (8/8; chance 0.25 per case). Alternative intensity encodings and their biophysical implementations are compared in **Online Resource 1, Sect. S7**.

Spike economy. Wave coding requires only one input spike per active site: a 3×3 pattern therefore uses nine input spikes, compared with 36–72 for the tested burst- and facilitation-based codes.

Compact read-out. The population response is summarized by a small fixed set of eight spatiotemporal moments. Computed directly from full-resolution 28×28 MNIST images, bypassing the network, these descriptors reached an accuracy of 0.63 with the fixed nearest-mean read-out and 0.79 with k-nearest-neighbour (kNN) or quadratic-discriminant (QDA) classifiers ($n = 3000$ held-out images; chance 0.10; **Online Resource 1, Sect. S3**). This characterizes the representational capacity of the feature space, not the performance of the WST-CA3 network.

3.9 Wave coding is dispensable when responses remain localized

The preceding experiments use travelling-wave input throughout. To determine whether the wave itself is necessary for categorization, we separate its effect from that of feedforward spatial localization. Both mechanisms can restrict the L-population response: a compact $I \rightarrow L$ projection does so anatomically, whereas a travelling wave can do so temporally even when the anatomical projection is broad. Guzman et al. (2021) also show that the spatial structure of connectivity is not merely an anatomical detail, but that it can determine the type of computation performed by a network.

A second unexpected result emerged when we tested whether travelling-wave coding itself is necessary for categorization (§2.10). With compact feedforward connectivity, simultaneous input achieved categorization comparable to wave-coded input (0.533 vs. 0.467). When connectivity was broadened, however, simultaneous input produced a much denser L-response than wave-coded input (64.5 vs. 26.6 firing neurons) and categorization fell to 0.200, no longer distinguishable from chance (**Fig. 8**). Wave-coded input, in contrast, maintained sparse responses and categorization of 0.467 at both footprint widths.

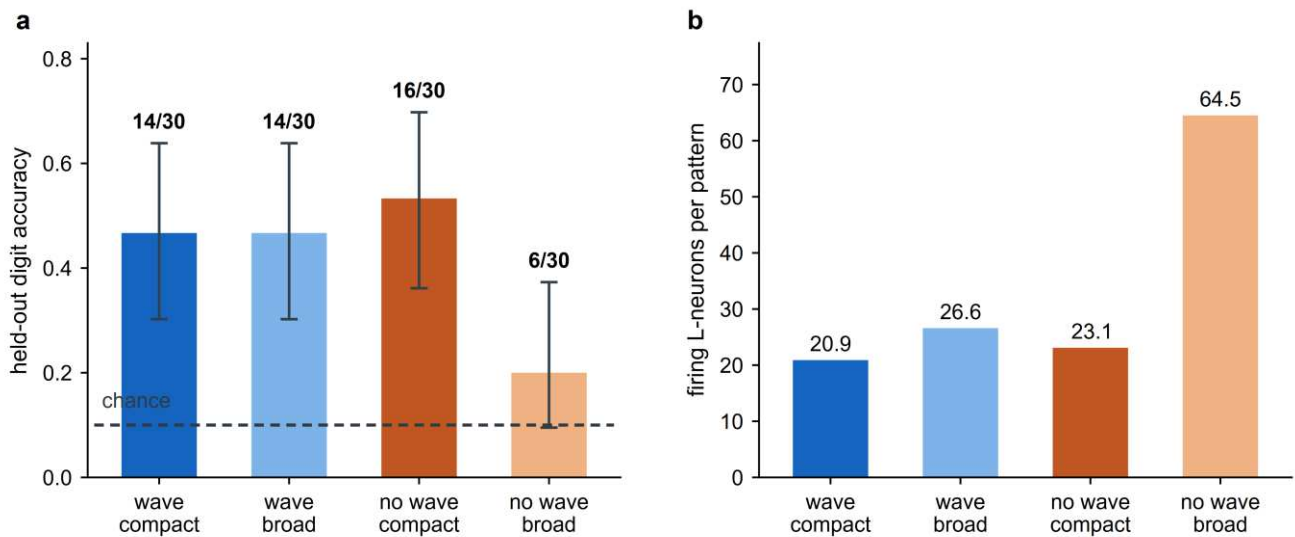


Fig. 8 Response localization rather than travelling-wave coding per se predicts categorization. (a) Held-out MNIST digit categorization accuracy and (b) the mean number of firing L-neurons per pattern, at a compact and a broad feedforward projection footprint, under wave-coded and simultaneous input. Bars in (a) are labelled with the number of correct classifications out of 30 held-out digits; whiskers are 95 % confidence intervals and the dashed line is the chance level of 0.10. With wave coding, categorization remains high and the population response stays sparse at both footprint widths. With simultaneous input, broad feedforward connectivity produces substantially increased L-neuron firing and categorization no longer distinguishable from chance ($P = 0.07$, exact binomial), whereas a compact footprint restores categorization despite the absence of the wave. With 30 digits per condition the intervals are wide and small differences between conditions are not resolvable; Online Resource 1, Sect. S11 gives the pairwise tests

Thus, successful categorization was associated with a spatially restricted L-population response rather than uniquely with travelling-wave timing. The wave becomes important when anatomical connectivity is broad, where it provides functional localization that simultaneous input does not; this role is operation-specific, since temporal pattern separation (§3.8) depends directly on wave-generated timing differences, whereas categorization can remain successful without the wave when spatial localization is supplied anatomically. A detailed analysis of the underlying space-to-time transformation — the full footprint experiment with its pairwise statistics and the sx-fsS mechanism — is provided in **Online Resource 1, Sect. S11**.

4. Discussion

4.1 Recognition before learning: how space is encoded in time

The central result of this study is that translation-invariant category structure can arise from fixed network dynamics before any synaptic learning occurs. In WST-CA3, this structure is generated by a network that exploits the temporal dimension for spatial pattern processing.

The network begins with a spatial input that is temporally unfolded along one dimension (x) by wave coding: travelling-wave activity converts spatial position into relative spike timing. Such waves are a well-established component of cortical dynamics at both local and large scales and can transiently modulate neuronal excitability and spiking as they propagate (Muller et al. 2018); in the hippocampus, theta oscillations themselves propagate as travelling waves along the septotemporal axis (Lubenov and Siapas 2009; Patel et al. 2012). Cortical waves have also been proposed to embed recent input history into the instantaneous population state (Muller et al. 2024); WST-CA3 uses a related principle within a single sweep.

The wave-encoded input reaches the L-neurons through dendritic branches with pathway-specific delays, and feedback inhibition acts with its own delay; recurrent $L \rightarrow L$ excitation is gated off during this early feedforward response (§2.8). The temporal extent created by wave coding, together with this delay structure, determines the first-spike times of the L-neurons. These first spikes form the output of WST-CA3: an L-spike cloud that codes space in time.

Our simulations show how this architecture uses the temporal dimension to generate an L-first-spike cloud with substantial potential for downstream processing:

- a) The L-cloud already contains category information without learning (§3.2). A fixed centroid-relative moment read-out extracts this information, and **Online Resource 1, Sect. S3** shows that it can be approximated by spiking feature neurons. Different tasks draw on different subsets of the same fixed feature set (§3.4), so categorization does not require a separately learned representation for each task: the centroid-relative feature space provides a low-dimensional basis from which different category boundaries can be recovered.
- b) The L-cloud can be read out in a translation-invariant manner. Spatial descriptors expressed relative to the centroid of the L-spike cloud are invariant to rigid shifts by construction; whether the network preserves the category-relevant geometry across positions is an empirical question. The held-out-position experiments answer it: category identity remains decodable at positions excluded from the class descriptors, whereas absolute position cannot be decoded above chance with the tested read-out (§3.3).
- c) The L-cloud is compatible with learning in the recurrent pathway. Categorization uses the early feedforward response, whereas pattern completion recruits learned recurrent $L \rightarrow L$ connectivity in a later retrieval phase, so the two operate in different temporal regimes of the same population (**Online Resource 1, Sect. S8**). This retains the classical CA3 recurrent-collateral principle of pattern completion within the substrate that supports categorization and temporal separation. The completion experiments are exploratory and establish compatibility rather than a fully characterized retrieval model.
- d) The L-cloud supports pattern separation. Spatially overlapping patterns become less overlapping in first-spike timing even when their spatial population overlap increases (§3.8). Removing the wave reverses this effect, showing that separation is carried specifically by wave-generated timing rather than by spatial

expansion. This extends our previous one-dimensional account (Löffler and Gupta 2022) to two-dimensional population activity. Spatial expansion and temporal coding thus provide distinct routes to separation: one enlarges the representational space through sparse recoding, the other exploits an additional temporal dimension within the same population.

e) A single sparse code can carry form, timing and intensity. Wave coding requires only one timing-bearing input spike per active site, while stimulus intensity can be represented by the strength of the corresponding synaptic event (§3.8). Spatial position, relative timing and intensity can therefore coexist within the same event-based representation rather than requiring separate coding pathways. Although demonstrated only for a small designed stimulus set, this result shows that adding intensity need not disrupt the spatial timing code.

f) Wave coding restricts the functional reach of broad projections along the wave axis. Because position along x is converted into input latency, only afferents arriving within the brief coincidence window summate, so a broad feedforward projection yields a response almost as localized as a compact one (26.6 versus 20.9 firing L-neurons, compared with 64.5 without the wave). When the projection itself is compact, the wave is not required for the category information of the L-cloud (§3.9).

These findings are unlikely to exhaust the potential of the temporal dimension for spatial pattern processing. Temporal coding has long been recognized as a central component of neural computation, and WST-CA3 combines it with established principles of pattern recognition — dendritic delays and coincidence detection, sparse competitive coding and moment-based invariant description — in a single model. The next section (§4.2) discusses where WST-CA3 aligns with or diverges from previous models.

4.2 Relation to other models

WST-CA3 combines mechanisms that individually have substantial precedent but are usually treated separately (**Table 3**; extended comparison in **Online Resource 1, Sect. S9**). Temporal-coding models and Liquid State Machines likewise exploit spike timing, but typically use a trained downstream decoder (Buonomano and Merzenich 1999; Maass et al. 2002; Zhang et al. 2025). Predictive-coding models, models based on spike-timing-dependent plasticity (STDP) and relational models instead acquire category-relevant representations through learning (Diehl and Cook 2015; Hazan et al. 2018; Whittington et al. 2020; Kawahara and Fujisawa 2025). WST-CA3 differs in asking how much useful structure can be generated by fixed connectivity and dynamics before such learning occurs.

The model is also related to classical hippocampal accounts of pattern separation and completion. Dentate-gyrus theories emphasize sparse spatial expansion (Marr 1971; O'Reilly and McClelland 1994; Rolls 2013), whereas WST-CA3 demonstrates separation in first-spike timing without expansion into a larger neuronal population. These mechanisms are complementary rather than mutually exclusive. Likewise, dendritic delay and coincidence detection provide established primitives for temporally structured integration (Stuart and Häusser 2001; Calixto et al. 2008; Branco et al. 2010), used elsewhere for geometric feature detection, place-context binding and neuromorphic delay hardware (George et al. 2021; Katz et al. 2007; D'Agostino et al. 2024); WST-CA3 combines these with timing generated systematically by a travelling wave. Such sequence-sensitive convergence does not require a learned wiring diagram: it arises even under random connectivity constrained by realistic cortical statistics (Somashekar and Bhalla 2025).

Finally, the spatial feature representation follows the general principle of moment-based invariant pattern description (Hu 1962; Flusser et al. 2009), but the descriptors are applied to a network-generated spike cloud rather than directly to an image and are extended by temporal features. The distinctive element of WST-CA3

is therefore not any single component, but their combination: structured timing, fixed network transformation, and a fixed low-dimensional feature basis.

Recent work supports several of the mechanisms combined in WST-CA3. A recent synthesis has proposed neural travelling waves as a substrate for spatiotemporal, dynamics-based computation, in which activity propagating across topographic sensory maps creates dependencies between position and time that are not naturally captured by purely feedforward or feedback processing (Muller et al. 2026). WST-CA3 provides a specific instance of this more general principle: the travelling wave systematically converts spatial position into relative spike timing, making spatial relationships available to temporal network dynamics. The present results suggest that such a transformation can support not only temporal context and prediction, but also translation-invariant categorization and temporal pattern separation. Recent human recordings provide particularly direct support for temporal population coding in visual categorization: relative firing sequences within brief population bursts carried information about both visual categories and individual exemplars, beyond that conveyed by firing rates or stimulus-locked latencies (Xie et al. 2024). WST-CA3 provides a potential mechanism for generating such category-relevant temporal structure by transforming spatial relationships into relative spike timing through a travelling wave.

Table 3 Positioning of WST-CA3 relative to related models along six axes

Computational axis	Representative approaches	WST-CA3
Temporal coding and decoding; liquid state machine	Buonomano and Merzenich 1999; LSM (Maass et al. 2002; Zhang et al. 2025)	travelling-wave-generated timing; fixed feature read-out
Formation of category-relevant representation	predictive coding (Rao and Ballard 1999; Lee et al. 2024); Tolman–Eichenbaum Machine (Whittington et al. 2020; spiking form Kawahara and Fujisawa 2025); STDP networks (Diehl and Cook 2015; Hazan et al. 2018); untrained-network selectivity (Baek et al. 2021)	category-relevant structure present before synaptic learning
Pattern separation	DG expansion (Marr 1971; O'Reilly and McClelland 1994; Rolls 2013; Guzman et al. 2021)	temporal separation without population expansion
Temporal integration primitive	Stuart and Häusser 2001; Calixto et al. 2008; Branco et al. 2010; Somashekar and Bhalla 2025; Stoll et al. 2024; George et al. 2021; Katz et al. 2007; DenRAM (D'Agostino et al. 2024)	wave- generated temporal order + fixed coincidence-sensitive circuitry
Invariant feature basis	moment-based invariants (Hu 1962; Flusser et al. 2009); invariance emerging from random feedforward projections in untrained networks (Cheon et al. 2022)	centroid-relative spike-cloud descriptors, analytically specified and computed by spiking circuits, extended by first-spike timing
Low-spike coding	time-to-first-spike networks (Stanojevic et al. 2024)	single-spike input code, no training of the WST-CA3 transformation or read-out

4.3 The efficiency of not learning

If useful category structure is supplied by fixed circuit architecture and dynamics, the system avoids one specific computational cost: acquiring that transformation through repeated stimulus presentation and synaptic modification. In WST-CA3, the transformation examined here operates without task-specific training of the WST-CA3 network or fixed feature read-out, and each active input site contributes only a single timing-bearing spike — as in time-to-first-spike networks, which reach comparable sparsity but acquire it through training (Stanojevic et al. 2024). The model therefore addresses representational and coding efficiency rather than competition with trained classifiers in raw classification accuracy.

This does not make the computation cost-free; it shifts part of the burden from individual learning to architecture.

The broader possibility that useful selectivity can precede learning is not unique to WST-CA3. In rate-based deep networks, units selective to faces and to various non-face objects emerge robustly in randomly initialized, entirely untrained networks and suffice for face detection (Baek et al. 2021; **Table 3**); in the same class of networks that selectivity is invariant to viewpoint rotation, translation and scaling, an invariance traced to combinations of non-invariant units under random feedforward projections (Cheon et al. 2022). WST-CA3 reaches a comparable conclusion by the opposite route. There, invariance is an emergent statistical property of random high-dimensional projections; here it is constructed and analytically specified — the read-out is centroid-relative by definition, and absolute position is correspondingly not decodable from it (§3.3). WST-CA3 explores this principle in a spiking system: category-relevant structure arises from fixed connectivity interacting with spatiotemporal population dynamics rather than from task-trained synaptic weights. The mechanisms differ, but both results challenge the assumption that useful selectivity must always be acquired through experience-dependent optimization.

Architecture-based and learned computation therefore occupy different positions on an efficiency–flexibility trade-off. Fixed circuitry can make useful transformations immediately available, but only for computations supported by its predefined dynamics and read-out; learning incurs an acquisition cost but provides the flexibility to adapt representations and decision boundaries to novel categories and changing environments. The two strategies need not be alternatives: a fixed architecture may provide an initially structured representation on which subsequent learning can operate, reducing what must be acquired from experience (§4.5).

4.4 A possible link to interference-based representations

The combination of travelling-wave coding and recurrent pattern completion (§4.1; **Online Resource 1, Sect. S8**) suggests a more speculative connection to interference-based representations. Holographic memory has long provided an analogy for distributed, content-addressable storage in which information is represented through interference relations and can be reconstructed from partial input (Pribram 1991). WST-CA3 does not implement optical holography, a Fourier transform or a field-based memory mechanism, and the present simulations do not demonstrate holographic storage. The narrower analogy is computational: travelling-wave propagation converts spatial layout into structured relative spike timing, which fixed-delay, coincidence-sensitive circuitry can distinguish.

Our earlier work has developed both directions of this mapping: the coding of spatial structure by relative spike timing (Löffler et al. 2024), and the converse time → space transformation, in which phase-shifted

subthreshold oscillations map temporal spike patterns onto distributed spatial activity (Löffler and Gupta 2020; Löffler and Bahmer 2026). Taken together with the present results, these mechanisms suggest the possibility of a bidirectional space $\leftrightarrow$ time codec, extending this temporal-coding principle. This bidirectionality is at present a conceptual synthesis rather than a demonstrated closed-loop encoding-and-retrieval system.

A direct test would require an additional mechanism not present in the current WST-CA3 model. One possibility is a second, internally generated reference-like wave whose relative timing with stimulus-driven activity could be encoded in recurrent connectivity and reinstated during retrieval. The resulting model could then test whether partial input reconstructs distributed activity through stored timing relations, rather than assuming an analogy to holographic interference. Such an extension would turn the present conceptual link into a falsifiable computational hypothesis.

4.5 Limitations, outlook, and predictions

Several limitations constrain the present conclusions.

- The moment-feature representation is deliberately low-dimensional and therefore cannot preserve all stimulus information: with the network bypassed and the moments computed directly from full-resolution MNIST images, the fixed read-out reaches 0.63 and stronger classifiers on the same feature vectors 0.79 (§3.8) — bounds on the feature space, not estimates of network performance.
- The invariance established here is translation only: orientation is treated as a class distinction rather than a nuisance variable, and rotation and scale enter merely as small within-class jitter for the printed letters (§2.3). Broader transformation invariance has not been tested.
- The simulated networks are small, and the feature circuits are predefined rather than developmentally acquired.
- Intensity coding has been tested only in restricted conditions.
- The dissociation between wave timing and anatomical localization has so far been examined only for digit categorization, with 30 held-out digits per condition and a single network realization (§3.9).
- The completion pathway remains exploratory.
- Wave speed is a further open point: cortical and hippocampal travelling waves propagate considerably faster than the one-spike-per-step wave used here (Muller et al. 2026), so what matters computationally is the ratio of the per-column delay to the dendritic coincidence window rather than the absolute velocity.

A further limitation concerns the reach of the phrase "before learning". The fixed architecture produces category-relevant *structure*: same-class stimuli map to neighbouring points of the moment space and different classes to separated regions, and this requires no synaptic change. Assigning a *discrete category* to a new stimulus is a further step, and in the present work it is taken in one of two ways: the nearest-prototype read-out averages a small labelled reference set into class prototypes (§2.5), and the competitive layer adapts its templates by unsupervised Hebbian plasticity (**Online Resource 1, Sect. S4**). Neither modifies the WST-CA3 network or the feature representation, so the central claim is unaffected — but neither is obtained from wiring alone. Competition without adaptation would in principle suffice: a winner-take-all layer with fixed templates partitions the moment space into stable regions and returns one winner per stimulus, with no plasticity of any kind. Such a partition does align with the stimulus classes to a substantial degree: with fixed random templates that have never been exposed to a stimulus, the competitive layer of **Online Resource 1, Sect. S4** still assigns held-out presentations to orientation-specific cells well above chance, at roughly three quarters of the level reached after Hebbian adaptation. What plasticity adds, at least here, is the alignment of

an already class-respecting partition rather than the partition itself. How far this carries beyond the four orientation classes examined here, and how such a partition would be specified without the feature standardization used here, remain open.

An important open question is therefore how the predefined feature circuitry might arise biologically. WST-CA3 need not imply that the nervous system explicitly specifies neurons for statistical moments. The analytical features could instead describe outputs assembled from more elementary developmental primitives such as topographic convergence, weighted pooling, opponent organization, coincidence detection and normalization. Such circuits might arise evolutionarily or through unsupervised developmental calibration rather than category-specific learning.

The model makes several experimentally testable predictions. Early population responses should contain stimulus-geometry information in relative spike timing beyond firing rate and absolute latency; human single-unit recordings in which spike order within population bursts carries category information independently of rate and latency are consistent with this prediction (Xie et al. 2024). More specifically, the relationship between stimulus orientation and spike synchrony should depend on wave direction: rotating the travelling wave should rotate the orientation dependence of synchrony. Finally, disrupting travelling-wave timing should affect computations differently depending on anatomical convergence: temporal pattern separation should be directly impaired, whereas categorization may remain possible where feedforward connectivity itself provides sufficient spatial localization.

Overall, WST-CA3 suggests that time can provide an additional computational coordinate through which fixed neural dynamics expose category-relevant structure before learning.

5. Conclusion

WST-CA3 demonstrates that translation-invariant category information can emerge from fixed spiking-network dynamics before task-specific learning. A travelling excitability wave converts spatial structure into relative spike timing, while fixed network dynamics generate population responses from which category-relevant information can be extracted by a low-dimensional, centroid-relative spatiotemporal feature representation. Categorization remains possible at previously unseen positions, while absolute position is removed from the read-out.

The results also clarify the role of the travelling wave. Wave coding is not required for categorization when anatomical connectivity itself produces sufficiently localized population responses, but it provides functional localization under broad convergence and enables temporal pattern separation without expansion into a larger neuronal population. Thus, its computational contribution extends beyond translation invariance alone.

WST-CA3 therefore suggests a complementary relationship between architecture and learning: fixed circuit dynamics can provide an initially structured representation on which subsequent learning may operate, rather than requiring all category-relevant structure to be acquired through experience.

Statements and Declarations

Funding. No funding was received for conducting this study.

Competing interests. The authors have no competing interests to disclose.

Ethics approval. Not applicable. This is a purely computational study; it involved no human participants, no animal subjects and no personal data.

Consent to participate and consent for publication. Not applicable.

Author contributions. Both authors contributed to the conception of the study. The WST framework on which the present network builds was developed jointly by both authors (Löffler and Bahmer 2026). Hubert Löffler designed the WST-CA3 network, implemented the simulation and analysis software, carried out the simulations, analysed and interpreted the results, prepared the figures and wrote the first draft of the manuscript. Andreas Bahmer contributed to the conceptual framing of the study, critically evaluated the analyses and their interpretation, and supervised the project. Both authors revised the manuscript, and both read and approved the final version.

Data and code availability. The study generates no experimental data. All results are produced by simulation and are reproducible from the model implementation together with the pinned parameter set of **Online Resource 1, Table S6** (`config_4_01.py`), which fixes every parameter of the canonical configuration; per-experiment deviations are stated in the last column of that table. The canonical script for each reported result is listed in **Online Resource 1, Sect. S6**. Presentations use seeded jitter (`numpy.default_rng`), so a given seed reproduces identical spike times.

The following machine-readable result files accompany the code: `ablation_lopo_results.json` and the feature matrices `ablation_lopo*_seed7.npz` (§3.4); `seeds_3_7_3_8.json` (§3.7, §3.8); `consol_3_8_separation.json` (§3.8); `consol_3_7_circuits.json` (the §3.7 spiking-circuit accuracies, seeds 7/11/13); `consol_completion.json` (pattern completion, **Online Resource 1, Sect. S8**, seeds 7/11/13); `wave_footprint_accuracy.json` and `sx_vs_fsS_digits_data.json` (§3.9); `explain_dump_mnist3.json` (the worked example of **Online Resource 1, Figure S3**); `consol_3_2_wave_control.json` (§3.2 raw separation across seeds and without the travelling wave, **Online Resource 1, Sect. S10**); `check_3_6_gate.json` (the §3.6 gating controls); `check_branch_mapping.json` (the branch-label permutation test, **Online Resource 1, Sect. S1**); and `paper_numbers.json`, which holds the values drawn by the remaining Results figures. Individual quantities in §3.8 were recorded from console output rather than written to a result file; the runs that produced them are listed with their scripts and dates in the repository run log, and re-running the named script regenerates them.

The handwritten digits are the publicly available MNIST test set (LeCun et al. 1998), downsampled to 10×10 as described in §2.9. The printed letters A–J are generated by the code from the Arial Bold system font at 22 pt; a different font changes the stimuli, so the font is part of the specification for reproducing §3.5.

The model implementation, all experiment scripts, the pinned configuration and the result files listed above are openly available in the Zenodo repository at <https://doi.org/10.5281/zenodo.22899664>, released under the MIT licence. The deposit does not redistribute the MNIST test set or the Arial Bold font; both are obtained as described above and in the repository README.

References

- Baek S, Song M, Jang J, Kim G, Paik S-B (2021) Face detection in untrained deep neural networks. *Nat Commun* 12:7328. <https://doi.org/10.1038/s41467-021-27606-9>
- Branco T, Clark BA, Häusser M (2010) Dendritic discrimination of temporal input sequences in cortical neurons. *Science* 329:1671–1675. <https://doi.org/10.1126/science.1189664>
- Buonomano DV, Merzenich MM (1999) A neural network model of temporal code generation and position-invariant pattern recognition. *Neural Comput* 11:103–116. <https://doi.org/10.1162/089976699300016836>
- Calixto E, Galván EJ, Card JP, Barrionuevo G (2008) Coincidence detection of convergent perforant path and mossy fibre inputs by CA3 interneurons. *J Physiol* 586:2695–2712. <https://doi.org/10.1113/jphysiol.2008.152751>
- Cheon J, Baek S, Paik S-B (2022) Invariance of object detection in untrained deep neural networks. *Front Comput Neurosci* 16:1030707. <https://doi.org/10.3389/fncom.2022.1030707>
- D'Agostino S, Moro F, Torchet T, Demirağ Y, Grenouillet L, Castellani N, Indiveri G, Vianello E, Payvand M (2024) DenRAM: neuromorphic dendritic architecture with RRAM for efficient temporal processing with delays. *Nat Commun* 15:3446. <https://doi.org/10.1038/s41467-024-47764-w>
- Diehl PU, Cook M (2015) Unsupervised learning of digit recognition using spike-timing-dependent plasticity. *Front Comput Neurosci* 9:99. <https://doi.org/10.3389/fncom.2015.00099>
- Flusser J, Suk T, Zitová B (2009) Moments and moment invariants in pattern recognition. Wiley, Chichester. <https://doi.org/10.1002/9780470684757>
- George JK, Soci C, Miscuglio M, Sorger VJ (2021) Symmetry perception with spiking neural networks. *Sci Rep* 11:5776. <https://doi.org/10.1038/s41598-021-85232-3>
- Gollisch T, Meister M (2008) Rapid neural coding in the retina with relative spike latencies. *Science* 319:1108–1111. <https://doi.org/10.1126/science.1149639>
- Guzman SJ, Schlögl A, Espinoza C, Zhang X, Suter BA, Jonas P (2021) How connectivity rules and synaptic properties shape the efficacy of pattern separation in the entorhinal cortex–dentate gyrus–CA3 network. *Nat Comput Sci* 1:830–842. <https://doi.org/10.1038/s43588-021-00157-1>
- Hazan H, Saunders DJ, Sanghavi DT, Siegelmann H, Kozma R (2018) Unsupervised learning with self-organizing spiking neural networks. In: 2018 International Joint Conference on Neural Networks (IJCNN). IEEE, pp 1–6. <https://doi.org/10.1109/IJCNN.2018.8489673>
- Henze DA, Wittner L, Buzsáki G (2002) Single granule cells reliably discharge targets in the hippocampal CA3 network in vivo. *Nat Neurosci* 5:790–795. <https://doi.org/10.1038/nn887>
- Hopfield JJ (1995) Pattern recognition computation using action potential timing for stimulus representation. *Nature* 376:33–36. <https://doi.org/10.1038/376033a0>
- Hu M-K (1962) Visual pattern recognition by moment invariants. *IRE Trans Inf Theory* 8:179–187. <https://doi.org/10.1109/TIT.1962.1057692>

Katz Y, Kath WL, Spruston N, Hasselmo ME (2007) Coincidence detection of place and temporal context in a network model of spiking hippocampal neurons. *PLoS Comput Biol* 3:e234. <https://doi.org/10.1371/journal.pcbi.0030234>

Kawahara D, Fujisawa S (2025) The Spiking Tolman–Eichenbaum Machine: emergent spatial and temporal coding through spiking network dynamics. *bioRxiv*. <https://doi.org/10.1101/2025.10.16.682754>

LeCun Y, Bottou L, Bengio Y, Haffner P (1998) Gradient-based learning applied to document recognition. *Proc IEEE* 86:2278–2324. <https://doi.org/10.1109/5.726791>

Lee K, Dora S, Mejias JF, Bohte SM, Pennartz CMA (2024) Predictive coding with spiking neurons and feedforward gist signaling. *Front Comput Neurosci* 18:1338280. <https://doi.org/10.3389/fncom.2024.1338280>

Lim C, Blume HW, Madsen JR, Saper CB (1997) Connections of the hippocampal formation in humans: I. The mossy fiber pathway. *J Comp Neurol* 385:325–351. [https://doi.org/10.1002/\(SICI\)1096-9861\(19970901\)385:3%3C325::AID-CNE1%3E3.0.CO;2-5](https://doi.org/10.1002/(SICI)1096-9861(19970901)385:3%3C325::AID-CNE1%3E3.0.CO;2-5)

Löffler H, Bahmer A (2026) The wave-coded brain: transforming spike patterns through space-time dynamics. *Biol Cybern* 120:9. <https://doi.org/10.1007/s00422-026-01034-8>

Löffler H, Gupta DS (2020) A model of memory linking time to space. *Front Comput Neurosci* 14:60. <https://doi.org/10.3389/fncom.2020.00060>

Löffler H, Gupta DS (2022) A model of pattern separation by single neurons. *Front Comput Neurosci* 16:858353. <https://doi.org/10.3389/fncom.2022.858353>

Löffler H, Gupta DS, Bahmer A (2024) Neural coding of space by time. *Biol Cybern* 118(3–4):215–227. <https://doi.org/10.1007/s00422-024-00992-1>

Lubenov EV, Siapas AG (2009) Hippocampal theta oscillations are travelling waves. *Nature* 459:534–539. <https://doi.org/10.1038/nature08010>

Maass W, Natschläger T, Markram H (2002) Real-time computing without stable states: a new framework for neural computation based on perturbations. *Neural Comput* 14:2531–2560. <https://doi.org/10.1162/089976602760407955>

Madar AD, Ewell LA, Jones MV (2019) Temporal pattern separation in hippocampal neurons through multiplexed neural codes. *PLoS Comput Biol* 15:e1006932. <https://doi.org/10.1371/journal.pcbi.1006932>

Marr D (1971) Simple memory: a theory for archicortex. *Philos Trans R Soc Lond B Biol Sci* 262:23–81. <https://doi.org/10.1098/rstb.1971.0078>

Muller L, Busch AN, Davis ZW, Reynolds JH (2026) Neural traveling waves in cortex: network mechanisms and potential roles in neural computation. *Neuron*. <https://doi.org/10.1016/j.neuron.2026.06.019>

Muller L, Chavane F, Reynolds J, Sejnowski TJ (2018) Cortical travelling waves: mechanisms and computational principles. *Nat Rev Neurosci* 19:255–268. <https://doi.org/10.1038/nrn.2018.20>

Muller L, Churchland PS, Sejnowski TJ (2024) Transformers and cortical waves: encoders for pulling in context across time. *Trends Neurosci* 47:788–802. <https://doi.org/10.1016/j.tins.2024.08.006>

- Nicoll RA, Schmitz D (2005) Synaptic plasticity at hippocampal mossy fibre synapses. *Nat Rev Neurosci* 6:863–876. <https://doi.org/10.1038/nrn1786>
- O'Reilly RC, McClelland JL (1994) Hippocampal conjunctive encoding, storage, and recall: avoiding a trade-off. *Hippocampus* 4:661–682. <https://doi.org/10.1002/hipo.450040605>
- Patel J, Fujisawa S, Berényi A, Royer S, Buzsáki G (2012) Traveling theta waves along the entire septotemporal axis of the hippocampus. *Neuron* 75:410–417. <https://doi.org/10.1016/j.neuron.2012.07.015>
- Pribram KH (1991) *Brain and perception: holonomy and structure in figural processing*. Lawrence Erlbaum Associates, Hillsdale
- Rao RPN, Ballard DH (1999) Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nat Neurosci* 2:79–87. <https://doi.org/10.1038/4580>
- Rolls ET (2013) The mechanisms for pattern completion and pattern separation in the hippocampus. *Front Syst Neurosci* 7:74. <https://doi.org/10.3389/fnsys.2013.00074>
- Rolls ET (2016) Pattern separation, completion, and categorisation in the hippocampus and neocortex. *Neurobiol Learn Mem* 129:4–28. <https://doi.org/10.1016/j.nlm.2015.07.008>
- Rousseeuw PJ (1987) Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *J Comput Appl Math* 20:53–65. [https://doi.org/10.1016/0377-0427\(87\)90125-7](https://doi.org/10.1016/0377-0427(87)90125-7)
- Somashekar BP, Bhalla US (2025) Discriminating neural ensemble patterns through dendritic computations in randomly connected feedforward networks. *eLife* 13:e100664. <https://doi.org/10.7554/eLife.100664>
- Stanojevic A, Woźniak S, Bellec G, Cherubini G, Pantazi A, Gerstner W (2024) High-performance deep spiking neural networks with 0.3 spikes per neuron. *Nat Commun* 15:6793. <https://doi.org/10.1038/s41467-024-51110-5>
- Stoll A, Maier A, Krauss P, Gerum R, Schilling A (2024) Coincidence detection and integration behavior in spiking neural networks. *Cogn Neurodyn* 18:1753–1765. <https://doi.org/10.1007/s11571-023-10038-0>
- Stuart GJ, Häusser M (2001) Dendritic coincidence detection of EPSPs and action potentials. *Nat Neurosci* 4:63–71. <https://doi.org/10.1038/82910>
- Thorpe S, Gautrais J (1997) Rapid visual processing using spike asynchrony. In: *Advances in neural information processing systems* 9. MIT Press, Cambridge, pp 901–907
- Whittington JCR, Muller TH, Mark S, Chen G, Barry C, Burgess N, Behrens TEJ (2020) The Tolman–Eichenbaum Machine: unifying space and relational memory through generalization in the hippocampal formation. *Cell* 183:1249–1263. <https://doi.org/10.1016/j.cell.2020.10.024>
- Xie W, Wittig JH, Chapeton JJ, El-Kalliny M, Jackson SN, Inati SK, Zaghloul KA (2024) Neuronal sequences in population bursts encode information in human cortex. *Nature* 635:935–942. <https://doi.org/10.1038/s41586-024-08075-8>
- Zhang Y, Mo L, He X, Meng X (2025) Unsupervised spiking neural network based on liquid state machine and self-organizing map. *Neurocomputing* 620:129120. <https://doi.org/10.1016/j.neucom.2024.129120>