From Cortical Synchronous Rhythm to Brain-Inspired Learning Architecture: An Oscillatory Spiking Neural Network with Time-Delayed Coordination

Tingting Dan **Guorong Wu***
Departments of Computer Science and Psychiatry
University of North Carolina at Chapel Hill
Chapel Hill, NC 27599
{Tingting_Dan,grwu}@med.unc.edu

Abstract

Human cognition emerges from coordinated spiking dynamics in distributed neural circuits, where information is encoded by both firing rates and precise spike timing determined by brain rhythms. Inspired by this notion, we propose a brain-inspired learning primitive in which cognition-level neural synchrony emerges through the interactions between micro-scale dynamics of spiking neurons and a macro-scale mechanism of oscillatory synchronization. Specifically, we model each parcel (e.g., a cortical region or an image pixel) in the target system as a spiking neuron embedded in a predefined connectivity scaffold. Low-level information is encoded in a spatiotemporal domain, where neurons are selectively grouped and fire over time through self-organized dynamics. Oscillatory synchronization is driven by a latent projection of the input. Since brain dynamics operate in a regime of partial and transient synchronization rather than global phase locking, we model oscillatory coordination using a time-delayed synchronization formulation, which enables a top-down (macro-to-micro) modulation of heterogeneous neural spiking for a large-scale distributed system. Together, we devise a spiking-by-synchronization neural network (S^2-Net) that uses rhythmic timing as a control mechanism for efficient information processing. Promising results have been achieved across a broad range of tasks, including neural activity decoding, disease diagnosis, temporal binding and temporal signal processing.

1 Introduction

Grouping distributed features into coherent entities, i.e., the binding problem, remains a central challenge for neural networks (Greff et al., 2020). Object-centric models address it with slot-based competition (Locatello et al., 2020; Seitzer et al., 2023) or with complex-valued and rotating representations whose phases encode grouping (Löwe et al., 2022; Stanić et al., 2023; Löwe et al., 2023). In the brain, rhythmic coordination is thought to support selective communication among neural populations (Fries, 2005, 2015), although whether binding relies on synchrony per se or on object-based attentional gain remains debated (Roelfsema, 2023). We do not take a position on this debate. Instead, we use phase and synchrony as a computational inductive bias and ask whether an oscillator–spiking architecture built on it yields useful representations for graph-structured and sequential data.

Relevant works. A broad class of brain-inspired deep models is built on spiking neural networks (SNNs) and neuromorphic computing, which replace continuous activations with event-driven spikes

*Corresponding author.

40th Conference on Neural Information Processing Systems (NeurIPS 2026).

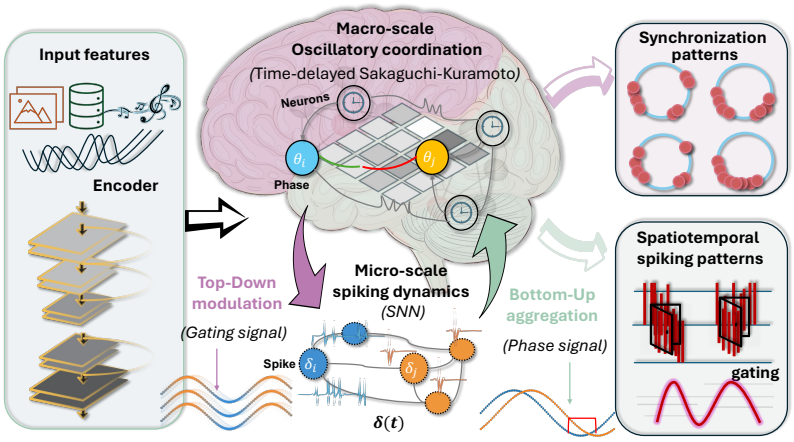

Figure 1: Overview of S^2-Net. The framework integrates a macro-scale oscillatory coordination module with a micro-scale spiking dynamics module. Delayed oscillatory phases rhythmically gate neural firing (macro-to-micro modulation). The phase lags of the oscillatory layer are derived from the connectivity scaffold, and the whole system is trained end-to-end through the spiking readout.

to capture temporal dynamics and improve energy efficiency (Maass, 1997; Roy et al., 2019). Recent advances in surrogate gradient learning have enabled SNNs to scale to vision and temporal tasks (Neftci et al., 2019). However, current SNNs have not yet matched the efficiency and robustness of biological neural systems on complex multiscale dynamical tasks. To address this limitation, Rhythm-SNN (Yan et al., 2025) uses heterogeneous oscillatory modulation to regulate spiking activity, enabling periodic activation at multiple frequencies. Another line of work draws inspiration from iterative inference and binding mechanisms in the brain. Predictive coding models introduce recurrent bottom-up and top-down updates to approximate inference in hierarchical architectures (Rao and Ballard, 1999). Closest to our work, Zheng et al. (2022) and Zheng et al. (2023) combine iterative bottom-up/top-down processing with spike-timing coherence to bind features through metastable dynamics in hybrid ANN–SNN models. In contrast, S^2-Net couples spiking neurons with a graph-constrained oscillator layer through phase lags and time delays. Reminiscent of oscillatory neural networks (Todri-Sanial et al., 2024), Kuramoto model (Kuramoto, 2005) is introduced to compress feature representations for the images (Miyato et al., 2025) and graphs (Dan et al., 2025; Ding et al., 2026) via the mechanism of oscillatory synchronization.

Although the Kuramoto model is a widely used framework for studying synchronization and has shown promise in brain-inspired machine learning, its standard formulation assumes instantaneous coupling and therefore favors zero-lag global synchronization. Such rigid phase locking has limitations from a perspective of dynamical system, as excessive synchrony collapses representational diversity and limits flexible information processing (Singer, 1999). In contrast, empirical evidence indicates that brain dynamics operate in a regime of partial and transient synchronization, where coordinated activity emerges selectively and dissolves as task demands change (Varela et al., 2001). Phase-lagged (Sakaguchi–Kuramoto) coupling and transmission delays both relax instantaneous coupling and give rise to metastable, non-zero-lag synchronization patterns (Yeung and Strogatz, 1999; Bronski et al., 2018; Cabral et al., 2022), which are a more suitable foundation for modeling brain-like coordination (Deco et al., 2017).

Our work. Cortical computation is often described as an interplay between stimulus-driven processing and contextual gain control that modulates it (Felleman and Van Essen, 1991; Rao and Ballard, 1999). We use this picture only as motivation. Rather than modeling a specific pathway, we abstract it into a two-layer computational architecture (Fig. 1) in which a population of coupled oscillators modulates a population of spiking neurons.

At the micro scale, spiking neurons encode information through sparse, temporally precise events. At the macro scale, oscillatory units regulate when each spiking neuron integrates its input (Fries, 2015). We model each parcel in the system, such as a cortical region or an image pixel, as a spiking neuron coupled with others in a predefined connectivity scaffold, so that low-level information is represented as spatiotemporal spike patterns. The oscillators follow a Sakaguchi–Kuramoto model whose phase lags are set by the connectivity scaffold, through an optimal-transport cost, and by a learned direction, which supports metastable, non-zero-lag synchrony. Their phases, read out with a time delay, rhythmically gate the membrane integration of the spiking neurons, and gradients of

the downstream loss flow back through these gates to the oscillator parameters. Together, these components form a spiking-by-synchronization neural network (S^2-Net).

Scope of the contribution. S^2-Net is an architecture, not a new learning rule: all parameters are trained end-to-end with backpropagation through time on a downstream loss. Its contribution is an inductive bias, namely a feature extractor whose representations are shaped by the interaction between oscillating and spiking units during the forward pass. We use phase and synchrony as a computational device and make no claim that the brain binds features through synchrony (Roelfsema, 2023). Likewise, when S^2-Net is applied to fMRI, its spikes and membrane-like states are latent model variables rather than estimates of neuronal activity. We evaluate S^2-Net on brain-state decoding and disease classification from six fMRI datasets, on proof-of-concept temporal binding, and on standard temporal benchmarks without a predefined graph.

Our major contributions are threefold:

- We unify SNN architectures and neural modulation within an oscillator–spiking framework, where a graph-constrained Sakaguchi–Kuramoto layer with OT-derived phase lags gates spiking integration through its delayed phases (top-down, macro-to-micro modulation). This design is inspired by the neuroscience notion that cognition emerges from the interaction between precise spike-based information flow and synchronization-driven temporal coordination.
- We combine a phase lag in the oscillator coupling with a time delay in the gating, which relaxes zero-lag synchronization. Ablations show that the phase lag, the structural prior and the delayed gating each contribute to performance.
- We show that the resulting representations improve brain-state decoding and disease classification on six neuroimaging datasets and transfer to temporal benchmarks without a predefined graph, with a compact model (0.22 MB).

2 Method

2.1 Network Architecture

Many data of interest, such as regional brain signals, audio frequency channels or image pixels, can be described as a set of interacting entities that evolve over time. This motivates an oscillation- and synchronization-based inductive bias for learning their representations.

2.1.1 Model input and network components

Following this notion, we propose a brain-inspired neural network architecture S^2-Net, with two main components. *First*, instead of treating the input as a vector or an image, we model a generic sensory input $X \in \mathbb{R}^{T\times N}$ as a dynamical system composed of N interacting entities (e.g., anatomical regions, frequency channels, image pixels or feature dimensions), each evolving as a time series over T time steps. Thus, unlike conventional artificial neural network (ANN) architectures that are explicitly tailored to specific data modalities (e.g., CNNs for images and RNNs for sequential data), our S^2-Net does not assume a specific data modality. Depending on the data domain, X may represent a time-series of neural activity, a sequence of audio frames, or a static image encoded as a temporal sequence. The interactions among entities are encoded by a weighted graph $A \in \mathbb{R}^{N\times N}$, instantiated from biological structural connectivity, spatial proximity, or inferred from data. Take functional neuroimages of human brain as an example, $X_i(t)$ denotes the BOLD (blood oxygenation level–dependent) activity of brain region i at time t, and A represents the inter-regional coupling strength based on fiber counts.

Second, the model backbone is an interactive spiking–synchronization (S^2) architecture in which representations emerge from the interaction between micro-scale spiking dynamics (by SNN) and a macro-scale mechanism of oscillatory coordination (by Sakaguchi-Kuramoto model). As shown in Fig. 1-middle, we deploy paired **spiking neurons** $\mathbb{S} = \{\mathscr{S}_1, \mathscr{S}_2, ..., \mathscr{S}_N\}$ at the bottom layer and corresponding **gating neurons** $\mathbb{G} = \{\mathscr{G}_1, \mathscr{G}_2, ..., \mathscr{G}_N\}$ at the upper layer, each pair is associated with the data entity X_i. The spiking activity on $\mathscr{S}_i$ is represented as a binary time series $\delta_i(t) \in \{0, 1\}$, where 0 and 1 denote the inactive and active states, respectively. In parallel, each gating neuron $\mathscr{G}_i$ maintains a phase state $\theta_i(t)$. The coupling between $\mathscr{G}_i$ and $\mathscr{G}_j$ carries a phase lag α_{ij} derived from the connectivity scaffold, so that connected gating neurons synchronize with a structured, non-zero

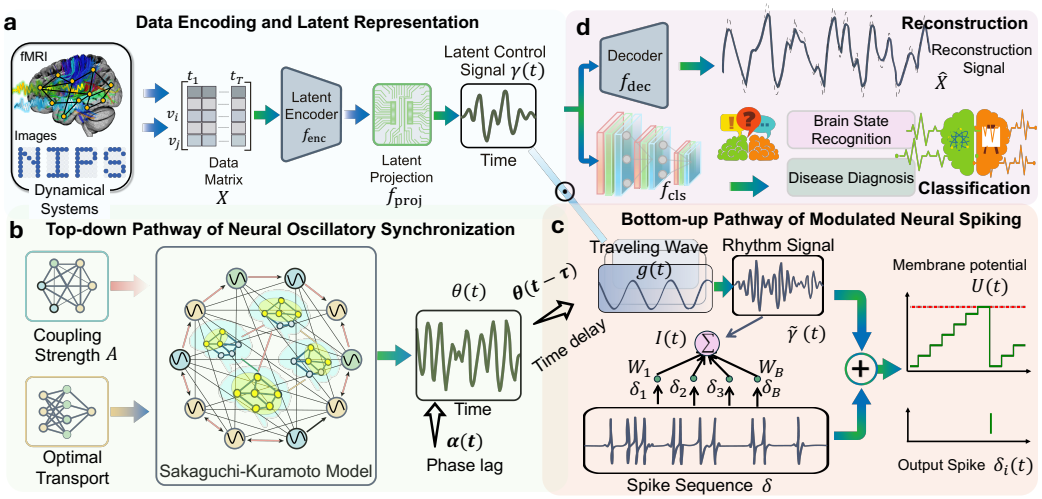

Figure 2: S^2-Net architecture. (a) Encoding inputs to latent signal $\gamma(t)$. (b) Sakaguchi-Kuramoto model governs phase states $\theta(t)$. (c) Delayed phases $\theta(t-\tau)$ gate $\gamma(t)$ to modulate membrane potential $U(t)$ and drive SNN spike output. (d) Decoding and classification for downstream tasks.

phase offset. Meanwhile, we introduce a time delay among gating neurons to induce a rhythm signal for modulating spiking patterns across spiking population $\mathbb{S}$. Together, phase lag and time delay couple micro-scale spiking dynamics with macro-scale oscillatory coordination. Unlike delay-free coupling, this combination allows partial synchronization and directional interactions among units.

2.1.2 Model components and training

S^2-Net comprises the following components. All learnable parameters are fitted jointly by backpropagation through time (Algorithm 1).

Latent projection of control signals. To extract coordination-relevant structure from high-dimensional inputs, we employ a two-stage encoding strategy to map raw signals into a latent driving representation (Fig. 2a):

$$\gamma(t) = f_{\text{proj}}(z(t)) = f_{\text{proj}}(f_{\text{enc}}(X(t))), \gamma(t) \in \mathbb{R}^N. \tag{1}$$

First, we deploy an encoder f_{enc} to remove external noise from the input signal $X(t)$. *Second*, we project the noise-free signals $z(t)$ to a time-varying *control signal* $\gamma(t)$, which modulates the neural oscillatory synchronization dynamics. Specifically, $\gamma(t)$ acts on the vector-valued *phase state* $\Theta(t) = \{\theta_i(t) | \theta_i(t) \in \mathbb{R}^D, i = 1, ..., N\}$ and steers its evolution under the Kuramoto dynamics. This design is inspired by the vector-valued Kuramoto dynamics (Miyato et al., 2025; Dan et al., 2025), enabling the system to "bind" desired task-relevant patterns through synchronization.

Overview of the coordination scheme. The paired gating and spiking neurons abstract the interplay between input-driven activity and contextual gain control. Their interaction involves: (i) controlled neural oscillatory synchronization with phase lags for temporal binding and segregation, (ii) rhythmic gating for temporal multiplexing from traveling waves, and (iii) rhythm-modulated spiking neural computation for downstream decoding. The working mechanism of each component is explained in Sec. 2.2

Training. Within a forward pass, the oscillator layer is integrated over $t = 1, \ldots, T$ and its delayed phases gate the spiking layer at every time step (Algorithm 1). The resulting spike-based features feed a task-specific readout, and all parameters are learned end-to-end by backpropagation through time with surrogate gradients (Neftci et al., 2019); we do not propose a new synaptic update rule. The dynamics of neural firing and oscillatory synchronization in S^2-Net also support model interpretation.

Remark. Throughout this paper, the term "learning architecture" refers to the computational scheme by which the forward dynamics of S^2-Net organize inputs into task representations: a macro-to-micro rhythmic modulation of spiking units by the delayed oscillatory phase state, together with end-to-end training through which the downstream loss on the spiking readout shapes the oscillator parameters.

Algorithm 1 Forward pass and training of S^2-Net

Require: input $X \in \mathbb{R}^{T\times N}$, connectivity A, time delay τ, task label y

Compute phase lags α_{ij} from A and the learned direction — Eq. (7)
Initialize phases $\theta_i(0)$, membrane-like states $U_i(0)$ and spikes $\delta_i(0)$
for $t = 1, \ldots, T$ **do**
  $\gamma(t) \leftarrow f_{\text{proj}}(f_{\text{enc}}(X(t)))$ — Eq. (1)
  Update phases $\theta(t)$ — Eq. (2)
  Compute gates $g_i(t)$ from $\theta_i(t-\tau)$ and rhythm signals $\tilde{\gamma}_i(t)$ — macro-to-micro
  Update dendritic currents, $U(t)$ and spikes $\delta(t)$ — Eq. (6)
  Read out $\mathcal{H}(t)$ from $\delta(t)$
end for
Compute the task loss on $\{\mathcal{H}(t)\}$ (Appendix A.3) and update all parameters by BPTT

2.2 Synchronized Spiking Architecture

2.2.1 Oscillatory synchronization and rhythmic gating

The oscillatory layer (Fig. 2b) updates the phase state of each gating neuron in the upper layer and generates the rhythm signals that selectively gate bottom-layer spiking neurons as a consequence of oscillatory phase synchronization. We formulate the evolution of vector-valued oscillatory state $\theta_i(t) \in \mathbb{R}^D$ in a topologically constrained Sakaguchi-Kuramoto system, which incorporates intrinsic frequencies, sensory forcing, and phase lags derived from the connectivity scaffold.

Coupled vector Kuramoto dynamics with time-dependent phase lags. For the d-th component in θ_i ($d \in \{1, \ldots, D\}$), the evolution of θ_i^d are determined by:

$$\dot{\theta}_i^d(t) = \omega_i^d + \kappa_i^d \sin\big(\gamma_i(t) - \theta_i^d(t)\big) + \frac{K}{N}\sum_{j=1}^{N} A_{ij} \sin\left(\theta_j^d(t) - \theta_i^d(t) - \alpha_{ij}\right) \tag{2}$$

where K is the global coupling strength, A_{ij} is the coupling strength constraining the interaction range, and α_{ij} characterizes the phase lag between two coupled gating neurons (Eq. 7). ω_i^d represents the learnable *natural frequency* of the oscillator, capturing the intrinsic characteristic of neural vibration. κ_i^d controls the strength of the attraction toward the data-specific control pattern $\gamma_i(t)$, ensuring that the dynamical system reacts to the incoming external stimuli.

Intuition of oscillatory synchronization in Sakaguchi-Kuramoto model. The Sakaguchi–Kuramoto model in Eq. (2) shapes oscillatory synchronization by phase-shifted interactions rather than instantaneous alignment. $\gamma_i(t)$ in the second term is an external "push" on the phase state that steers the synchronization pattern toward a desired phase configuration. Thus, the dynamics in Eq. (2) can settle into metastable phase-locked patterns, which then serve as task-relevant rhythmic gating signals. The third term in Eq. 2 introduces a phase lag α_{ij} derived from the connectivity scaffold. In the presence of this phase lag, the dynamics in Eq. 2 seek to align the phase-shifted oscillations, thereby generating rhythmic signals that selectively trigger bottom-layer spiking neurons.

Energy interpretation. Under the assumptions of symmetric coupling ($A_{ij} = A_{ji}$), antisymmetric phase lags ($\alpha_{ij} = -\alpha_{ji}$) and $\omega = 0$, the coupling and drive terms of Eq. (2) derive from the potential energy $V(\Theta, t)$:

$$V(\Theta, t) = \underbrace{-\frac{K}{2N}\sum_{d=1}^{D}\sum_{i,j} A_{ij}\cos(\theta_j^d - \theta_i^d - \alpha_{ij})}_{V_{\text{sync}}:\text{System dynamics}} \underbrace{-\sum_{d=1}^{D}\sum_{i}\kappa_i^d \cos(\gamma_i(t) - \theta_i^d)}_{V_{\text{data}}:\text{System control}} \tag{3}$$

With antisymmetric lags, V_{sync} is minimized when the phase difference between connected units matches the lag ($\theta_j^d - \theta_i^d \approx \alpha_{ij}$) rather than zero. Its minima are therefore phase-gradient, traveling-wave-like configurations instead of global in-phase states (Bronski et al., 2018; Ju et al., 2014; Muller et al., 2018). In parallel, V_{data} acts as a potential well anchoring the state under the control of $\gamma_i(t)$. As shown in Appendix A.2, when $\omega = 0$ and γ varies slowly relative to the phase dynamics,

Eq. (2) is a gradient flow of V, i.e., $\dot{\theta}_i^d = -\partial V/\partial \theta_i^d$, and V is non-increasing along trajectories. Under these assumptions, the phases settle into configurations that trade off agreement with the connectivity-constrained lags against consistency with the input drive. With heterogeneous natural frequencies or fast-varying inputs, this is an interpretation rather than a convergence guarantee.

Generating rhythm signals for modulating neural spiking. A growing body of neurophysiological evidence shows that cortical and hippocampal activity is organized via traveling oscillatory waves, spanning theta to gamma bands, which are closely linked to perception, attention, working memory, and learning (Deco et al., 2009; Breakspear, 2017; Muller et al., 2018). Following this notion, we further introduce a fixed time delay τ (a hyperparameter shared by all gating neurons) to generate the rhythmic gating signal $g(t)$ by binding with the delayed phase state $\theta_i(t-\tau)$:

$$g_i(t) = \sigma_{\text{gate}}\Big(\sin\big(\bar{\theta}_i(t-\tau)\big)\Big), \quad \bar{\theta}_i = \frac{1}{D}\sum_{d=1}^{D}\theta_i^d, \quad i = 1, \ldots, N, \tag{4}$$

where $\sigma_{\text{gate}}(x) = (1+x)/2$ maps the sine wave to $[0,1]$, so that $g_i(t)$ is a scalar gate. The time delay τ couples the two layers: each spiking neuron is gated by the phase its oscillator had τ steps earlier, so the phase relations established by Eq. (2) are converted into a temporal order of gating windows. We exploit this property to dynamically determine the grouping of spiking neurons through the rhythm signal $\tilde{\gamma}_i(t)$:

$$\tilde{\gamma}_i(t) = g_i(t)\,\gamma_i(t). \tag{5}$$

Through this process, static phase separation is converted into temporal multiplexing, so that spiking neurons are grouped by phase-aligned timing rather than by spatial convergence.

2.2.2 Rhythm-modulated spiking and OT-derived phase lags

Rhythm-modulated spiking neural network. Unlike point-neuron formulations, each spiking neuron comprises multiple dendritic branches with distinct temporal filtering constants, reflecting the compartmentalized integration observed in biological dendrites (London and Häusser, 2005). Functionally, a single leaky compartment has a monotonically decaying impulse response, whereas a weighted sum of B branches with different decay rates can realize non-monotone kernels, such as delayed peaks or band-pass responses. This allows a neuron to model fast and slow temporal dependencies at the same time. We use $B = 4$. The workflow of S^2-Net consists of the following major steps.

(1) Spike generation conditioned on rhythm. The input is an aggregation of B dendritic branches. The raw activation $J_{i,b}(t)$ of the b-th branch incorporates the rhythm-modulated signal $\tilde{\gamma}_i(t)$ on i^{th} spiking neuron as: $J_{i,b}(t) = W_{i,b}^{(\gamma)} \cdot \tilde{\gamma}_i(t) + \sum_j W_{ij,b}^{(\text{rec})}\delta_j(t-1)$, where $W_{i,b}^{(\gamma)}$ projects the oscillatory latent vector onto this selected branch (modulated by $\tilde{\gamma}_i(t)$), and $W_{ij,b}^{(\text{rec})}$ captures recurrent interactions. The branch state $l_{i,b}(t)$ evolves as a low-pass filter with a branch-specific decay rate $\beta_b \in (0,1)$: $l_{i,b}(t) = \beta_b l_{i,b}(t-1) + (1-\beta_b)J_{i,b}(t)$. Thus, the total dendritic current $I_i(t)$ is formed by aggregating the filtered outputs of all branches: $I_i(t) = \sum_{b=1}^{B} l_{i,b}(t)$.

(2) Modulated membrane potential update. The somatic membrane potential update is governed by the external rhythmic gate $g_i(t)$ derived in Eq. 4. We implement a soft gating mechanism where $g_i(t)$ modulates the effective integration rate of the neuron. The latent membrane-like state $U_i(t)$, an internal model variable rather than a measured membrane potential, is updated via interpolation between the previous state and the new leaky-integrated state $V_i(t)$:

$$V_i(t) = \rho\, U_i(t-1) + I_i(t), \quad U_i(t) = (1-g_i(t))\cdot U_i(t-1) + g_i(t)\cdot V_i(t) - \vartheta\delta_i(t-1), \tag{6}$$

where ρ is the membrane leak factor and ϑ is the firing threshold. This mechanism allows the rhythm to dynamically "freeze" ($g \to 0$) or "update" ($g \to 1$) the neural state, preserving information during off-cycles while enabling gradient propagation. The post-spike reset is applied outside the gate, so a neuron that fired at $t-1$ is reset even during an off-cycle and cannot fire repeatedly while its gate is closed. Pausing leak integration when $g_i(t)$ is low abstracts phase-dependent excitability and is loosely related to shunting inhibition in active cortical networks (Burman et al., 2023) and to activity-silent maintenance of information (De Pittà and Brunel, 2022). The output spike $\delta_i(t) \in \{0,1\}$ is generated as $\delta_i(t) = \Xi(U_i(t) - \vartheta)$, where $\Xi(\cdot)$ denotes a Heaviside step function.

(3) Spatiotemporal spiking feature aggregation. We employ a fully connected readout layer to map the population spiking activity $\delta(t)$ to a continuous-valued output $\mathcal{H}(t)$. This layer integrates

spatiotemporal spike patterns into a low-dimensional decision variable that is differentiable and amenable to optimization for downstream learning tasks.

OT-derived phase lags. The phase lag α_{ij} is not a free parameter. We use an *OT-derived phase lag*, i.e., a lag whose magnitude is given by the optimal-transport cost on the connectivity graph and whose sign is learned. Its magnitude is set by a transport cost on the connectivity graph, $C_{ij} \propto 1/(A_{ij} + \varepsilon)$, as used in graph optimal transport (Cuturi, 2013), min–max normalized to $[0, 1]$ and denoted $\tilde{C}_{ij}$. Its sign, i.e., which of the two units leads, is learned through an antisymmetric matrix:

$$\alpha_{ij} = \alpha_0 \tanh\big(W_{ij} - W_{ji}\big)\, \tilde{C}_{ij}, \tag{7}$$

where $W \in \mathbb{R}^{N\times N}$ is learnable and α_0 is a scale. Strongly connected pairs (low cost) are thus kept close to zero-lag synchrony, whereas weakly connected pairs can carry larger lags, which allows directed, traveling-wave-like coordination along the graph. Because A is symmetrized and $\tanh$ is odd, $\alpha_{ij} = -\alpha_{ji}$ holds by construction, which is the condition required by the energy interpretation in Sec. 2.2.1. The spiking layer does not feed back into the oscillator dynamics within a forward pass, instead, gradients of the task loss on the spiking readout propagate through the gates to W, the natural frequencies ω, the stiffness κ and the encoder.

Deploy S^2-Net in downstream applications We drive feature learning using three downstream applications (Fig. 2d): (1) *self-supervised reconstruction* as a regularizer, (2) *dynamic state recognition* for sequence labeling, and (3) *subject-level classification* via temporal pooling. Detailed formulations and loss functions are provided in Appendix A.3. The code is provided in GitHub.

2.3 Proof-of-Concept Simulation Design

Experimental setup. To illustrate *binding-by-synchrony* and *traveling-wave propagation*, we constructed two synthetic environments (Fig. 3): (a) disjoint squares and (b) nested topological structures[2]. Objects are initialized with distinct intensities (1.0 vs. 0.5) to enable feature-driven repulsion via the graph coupling term A_{ij}. We applied Gaussian perturbation at varying levels: a baseline $\sigma = 0.05$ for (a), and a severe noise regime $\sigma = 0.2$ for (b) to test robustness under low signal-to-noise ratio. Each of the $N = 28 \times 28 = 784$ pixels is one unit. In the spike rasters, the vertical axis is the flattened pixel index and the horizontal axis is the simulation step. Unlike S-MNIST, the image is not scanned as a sequence.

Discrete object discovery and attention rhythms (Fig. 3a). In the disjoint task (two objects shown in Fig. 3a), our S^2-Net breaks phase symmetry, driving the system from a disordered transient state into a stable anti-phase regime. Interestingly, the gating dynamics $g(t)$ (Fig. 3a-middle) emerge as smooth, alternating sinusoidal waves. This creates a "push-pull" mechanism that resembles rhythmic attentional sampling. The resulting spatiotemporal spike raster (Fig. 3a-right) reveals temporal chunking, where objects are phase-locked to alternating firing windows separated by refractory gaps.

Topological complexity and balanced competition (Fig. 3b). Standard coupled oscillators often collapse into in-phase synchronization or a winner-take-all regime where the larger ring suppresses the smaller square (Fig. 3b-left). In contrast, S^2-Net keeps both objects represented through graph-aware traveling waves. As shown in the gating signal map (Fig. 3b-middle), a phase gradient propagates along the ring, binding its spatially disjoint segments into a unified entity. As a result, the smaller square is not suppressed despite the size imbalance and competes equally for representation. The spike raster (Fig. 3b-right) reveals a traveling wave topology, where the square fires precisely within the local phase troughs of the ring's propagation, resolving the topological conflict through autonomous temporal multiplexing.

Quantitative evaluation and stability analysis. We evaluate feature binding using Kuramoto-derived metrics (Fig. 6d). The system rapidly transitions from a disordered state to a stable synchronized manifold. Specifically, intra-object synchronization converges to its maximum ($R_{intra} \to 1.0$), while inter-object overlap collapses ($R_{merge} \to 0$), consistent with anti-phase separation. See Appendix A.4.1 for detailed derivations. Because R_{merge} is also low for random phases and remains nonzero for perfectly separated objects of unequal size, it should be read together with R_{intra}.

[2]In computational neuroscience, nested structures are regarded as a canonical benchmark (Roelfsema et al., 1998; Chen, 1982) for figure-ground segregation. They challenge the system to solve the "binding problem" by integrating spatially separated segments into a coherent whole.

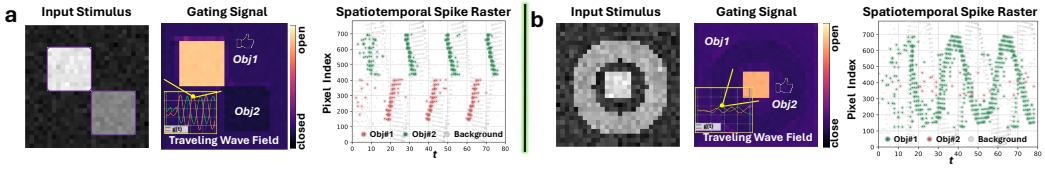

Figure 3: Proof-of-concept simulations of S^2-Net. (a) Disjoint Object Discovery. (b) Nested Topological Binding. Panels (left to right) show the input stimulus (28×28 pixels), the gating signals $g(t)$, and spatiotemporal spike rasters (vertical axis: flattened pixel index, 1–784, horizontal axis: simulation step).

Multi-object binding. On 32×32 scenes with 2–6 objects, we replace R_{merge} by size-invariant, chance-corrected metrics (FG-ARI, full-scene ARI and corrected coherence; Appendix A.4.1). Removing the phase lag halves FG-ARI ($0.28 \rightarrow 0.13$), and S^2-Net is close to AKOrN (Miyato et al., 2025) on FG-ARI (0.28 vs. 0.30) and higher on full-scene ARI (0.54 vs. 0.26).

3 Experiments

3.1 Experimental Setup

Comparison methods. We compare some representative SNN architectures, including: (i) a feedforward SNN (FFSNN) (Wu et al., 2019), (ii) a recurrent SNN (SRNN) (Zenke and Vogels, 2021), (iii) the long short-term memory SNN (LSNN) (Bellec et al., 2018), (iv) an adaptive SRNN (ASRNN) (Yin et al., 2020), (v) a dendritic heterogeneity SNN (DH-SNN) and (vi) Rhythm-SNN (Yan et al., 2025).

Human brain dataset. We evaluate S^2-Net on six large-scale neuroimaging datasets on both task recognition and disease diagnosis settings. For task recognition, we use HCP-A (Bookheimer et al., 2019), HCP-YA (Van Essen et al., 2013), and UK Biobank (UKB) (Miller et al., 2016) datasets, which include multiple task-based fMRI paradigms with corresponding diffusion MRI for constructing structural brain networks. In addition, we evaluate our method on the Working Memory task (HCP-WM) from HCP-YA, which involves alternating cognitive states within a single fMRI scan. This setting requires time-point–level task state recognition rather than assigning a single label to an entire scan. This formulation is more challenging, as it demands precise modeling of rapid task transitions, temporal dependencies, and subtle changes in functional brain dynamics. For disease diagnosis, we adopt ADNI (Jack Jr et al., 2008), PPMI (Marek et al., 2011), and NIFD (Planche et al., 2023) datasets, covering Alzheimer's disease (AD), Parkinson's disease (PD), and frontotemporal dementia (FTD), respectively, using resting-state fMRI and diffusion MRI data. Across all datasets, task recognition is formulated as a multi-class classification problem, whereas disease diagnosis is treated as a binary or multi-class classification task depending on the dataset. Detailed dataset descriptions and experimental settings are provided in Appendix A.1 and A.6. For subject classification, we use subject-level 10-fold cross-validation with the baseline scan of each subject. For quantitative evaluation, we use accuracy (Acc), precision (Pre), and F1-score (F1) to assess performance, while the quality of signal reconstruction is quantified using mean absolute error (MAE) and mean squared error (MSE).

3.2 Experiment Results on Neural Activity

Brain state recognition. Table 1 (left) presents the quantitative comparison of S^2-Net against representative SNN methods on HCP-A, HCP-YA, and UKB. The results demonstrate that S^2-Net consistently outperforms all baseline methods across all metrics ($^*p < 0.05$, paired-t-test). This holds even for DH-SNN and Rhythm-SNN, which incorporate dendritic heterogeneity and neural oscillations, respectively. This indicates that modeling the coupled interaction between structural and functional information allows for a more robust identification of distinct cognitive states across diverse subject populations.

Brain neural activity decoding. We further evaluate the model's capability in fine-grained temporal decoding using HCP-WM dataset. As shown in Table 1 (middle), this task is particularly challenging for conventional SNNs. For instance, FFSNN and simple recurrent models achieve much lower

Table 1: Performance comparison of S^2-Net against SNN models across human brain datasets via 10-fold cross-validation. '*' denotes significantly outperforming other methods ($p < 0.05$).

Model	Metric	HCPA (4 tasks)	HCP-YA (7 tasks)	UKB (2 tasks)	HCP-WM (9 tasks)	ADNI (3 states)	PPMI (4 states)	NIFD (5 states)
FFSNN	Acc	97.33±0.70	95.46±1.51	97.93±0.22	37.22±0.94	91.00±11.74	65.08±3.35	49.37±1.97
	Pre	97.03±0.94	95.62±1.50	97.94±0.22	35.26±0.91	94.50±7.14	**50.11±5.20**	38.00±7.14
	F1	95.99±1.03	95.47±1.50	97.93±0.22	33.07±0.88	89.27±13.72	35.41±6.19	29.63±3.40
SRNN	Acc	96.67±0.60	94.49±1.42	97.02±0.16	49.74±2.81	98.00±6.00	66.09±2.56	46.42±1.27
	Pre	95.53±1.11	94.61±1.35	97.02±0.16	47.66±3.28	98.75±3.75	45.75±4.66	35.18±11.07
	F1	94.89±0.99	94.47±1.41	97.02±0.16	45.73±3.19	97.62±7.14	39.84±4.60	25.35±4.12
LSNN	Acc	96.64±0.42	94.68±1.46	98.15±0.23	42.81±3.45	96.67±6.67	66.09±4.21	51.40±1.38
	Pre	95.77±0.90	94.78±1.39	98.17±0.23	40.00±3.50	97.75±4.53	47.21±5.01	43.34±9.14
	F1	94.81±0.79	94.63±1.49	98.15±0.23	38.60±3.77	96.06±7.95	38.46±7.39	35.90±3.30
ASRNN	Acc	96.46±0.59	92.83±1.25	98.79±0.08	78.07±3.11	96.33±7.37	64.38±3.78	46.06±2.38
	Pre	95.54±1.40	92.99±1.24	98.81±0.08	78.28±3.02	97.33±5.54	42.65±9.27	37.74±8.37
	F1	94.42±0.99	92.82±1.25	98.79±0.08	78.04±3.16	95.78±8.46	37.51±6.75	28.85±7.00
DH-SNN	Acc	97.92±0.49	95.66±1.96	98.58±0.15	56.64±1.34	98.00±6.00	65.40±4.87	47.73±1.35
	Pre	97.38±1.13	95.65±1.96	98.50±0.15	54.09±14.92	98.75±3.75	40.62±11.71	31.97±6.80
	F1	96.78±0.77	95.77±1.87	98.58±0.15	47.74±8.53	97.62±7.14	37.72±8.05	33.46±6.85
Rhythm-SNN	Acc	97.90±0.69	95.81±1.16	98.45±0.14	88.57±2.41	98.33±5.00	65.57±5.60	50.90±2.04
	Pre	97.10±1.34	95.79±1.17	98.47±0.14	88.37±2.64	**99.00±3.00**	41.20±11.72	43.38±7.75
	F1	96.78±1.18	95.88±1.18	98.45±0.14	88.23±2.70	97.78±6.67	36.72±9.06	39.75±2.36
S^2-Net	Acc	**98.97±0.07***	**96.98±0.03***	**99.72±0.14***	**88.72±2.09**	**98.33±5.00**	**67.62±5.21**	**64.65±1.79***
	Pre	**98.43±0.05***	**97.12±0.03***	**99.72±0.14***	**88.54±2.11**	98.75±3.75	49.30±8.83	**63.11±4.35***
	F1	**98.19±0.03***	**96.96±0.03***	**99.72±0.14***	**88.48±2.26***	**98.29±5.14***	**41.51±7.19***	**59.35±2.43***

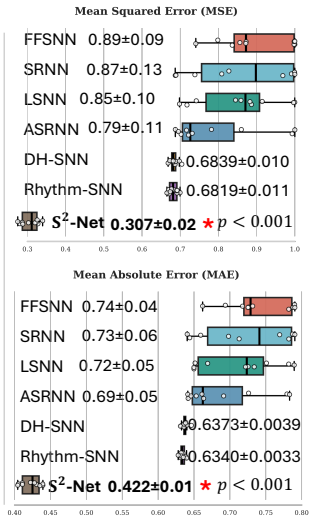

Figure 4: Brain decoding performance on HCP-WM dataset.

accuracy due to the complexity of the temporal dynamics. S^2-Net achieves the highest accuracy, on par with Rhythm-SNN (88.72% vs. 88.57%) and well above the other SNNs, suggesting that rhythmic gating helps decode rapid cognitive shifts. Furthermore, we assess the signal reconstruction quality to check that the learned representations retain information about the input signals rather than only optimizing for classification. Fig. 4 shows the reconstruction error on the HCP-WM dataset. S^2-Net achieves the lowest MSE and MAE among the compared methods, including Rhythm-SNN and DH-SNN. This low reconstruction error indicates that the latent states of S^2-Net retain the temporal structure of the input signals while supporting decoding.

Disease diagnosis. In the context of neurological disorder diagnosis (ADNI, PPMI, and NIFD), robustness to noise and small sample sizes is critical. As reported in Table 1 (right), S^2-Net obtains the highest or tied-highest accuracy among the compared SNNs for AD, PD, and FTD. While datasets like PPMI and NIFD present high inter-class similarity and limited training samples, our method obtains the highest accuracy and F1-score on both. This suggests that S^2-Net extracts discriminative features from these small, heterogeneous cohorts; whether these features have clinical value requires separate validation.

Latent firing patterns across neurodegenerative conditions. As a model-interpretation analysis, we inspect the latent spikes $\delta(t)$ of the trained classifiers (Fig. 5). These are model-derived events rather than measurements of neuronal firing, and the analysis is a proof of concept rather than a validation of biomarkers. In ADNI, the AD group shows a higher mean latent firing rate than CN, and the regions contributing most to classification lie in the medial temporal lobe and posterior hubs, which are commonly implicated in AD (Palop and Mucke, 2016). In NIFD, the global firing profiles of FTD and CN overlap, while the most discriminative regions cluster in frontal and temporal cortex. In PPMI, the PD group shows a faster rise and a higher steady-state latent firing rate than CN, with basal ganglia and sensorimotor regions ranked highest, in line with the circuits affected in PD (McGregor and Nelson, 2019). Whether these latent patterns relate to clinical or electrophysiological markers remains to be tested. More results are shown in Fig. 7 in the Appendix.

3.3 Versatility in General Temporal Processing

We next test whether S^2-Net generalizes to standard temporal benchmarks without a predefined graph. We adapt the S^2-Net to *learn* the underlying connectivity directly from the data. Specifically, we treat the adjacency matrix (and the associated transport lags) as learnable latent parameters, so that the model learns a synchronization topology for each task (Appendix A.2). We conduct several experiments across five widely used benchmarks: Sequential MNIST (S-MNIST) (Le et al., 2015), Permuted Sequential-MNIST (PS-MNIST) (Le et al., 2015), Spiking Heidelberg Digits (SHD)

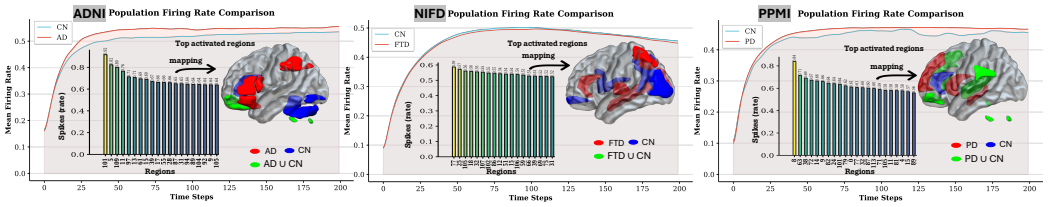

Figure 5: Latent spiking representations and the brain regions that contribute most to classification. There are three dataset-specific columns: ADNI, NIFD, and PPMI, arranged from global dynamics to regional patterns: Top Row. Mean firing rate curves for healthy control (CN, blue) and disease (red) groups. Bottom Rows. Ranked brain region bar charts based on classification contribution (quantified by spiking intensity (rates)) and their corresponding anatomical mapping to 3D brain projections.

(Cramer et al., 2020), Electrocardiogram (ECG) classification (Yin et al., 2021), and Google Speech Commands (GSC) (Warden, 2018) (Appendix A.1).

Table 2: Performance comparison (accuracy %) across datasets.

Model	S-M	PS-M	SHD	ECG	GSC
FFSNN	59.24	41.84	48.10	77.64	90.56
SRNN	81.51	64.89	81.60	71.56	91.76
LSNN	96.24	87.20	89.84	81.93	91.20
ASRNN	98.70	94.30	82.82	85.90	92.10
DH-SNN	98.90	94.52	89.84	86.35	94.05
Rhythm-SNN	98.94	96.73	91.07	86.41	94.47
S^2-Net	**98.97**	**96.94**	**91.39**	**89.23**	**94.58**

Discussion. In the main text, we solely compare our framework against SNN-series methods to compare within the spiking model family. To assess S^2-Net beyond the neuromorphic domain, we additionally compare it on the six neuroimaging datasets against general deep learning methods (e.g., LSTMs, Transformers, TCN, Mamba (Gu and Dao, 2023), and synchronization-based models, i.e., AKOrN (Miyato et al., 2025), GraphCON (Rusch et al., 2022), KuramotoGNN (Nguyen et al., 2024) and SR-Net (Harikrishnan and Nagaraj, 2021)). Please refer to Appendix A.4.2 for these detailed comparisons and full experimental results. The detailed ablation studies are shown in Appendix A.5.

Table 2 shows that S^2-Net transfers to tasks without a predefined graph. On PS-MNIST, which requires integrating information over 784 steps, it reaches 96.94%. It obtains the best accuracy among the compared SNNs on ECG (89.23%) and is competitive on SHD and on keyword spotting with GSC (94.58%). Margins over Rhythm-SNN on S-MNIST and GSC are small ($\leq$0.11 points) and are reported without variance, so we regard performance on these two tasks as comparable. These results suggest that the inductive bias is not specific to brain data.

4 Conclusion and Limitations

We presented the spiking-by-synchronization neural network (S^2-Net), a biologically inspired architecture in which a graph-constrained Sakaguchi–Kuramoto layer and a spiking layer interact within each forward pass: phase lags derived from the connectivity scaffold structure oscillatory coordination, and delayed oscillatory phases gate spiking integration. Across six neuroimaging datasets, S^2-Net matched or exceeded spiking baselines, and it transferred to temporal benchmarks without a predefined graph. Ablations show that the structural prior gives the largest gain, and that the phase lag and the delayed gating add further improvements.

Limitations. The spikes and membrane-like states of S^2-Net are latent model variables: on fMRI they evolve on the slow BOLD timescale and should not be read as neuronal spiking, and the disease-related firing patterns in Fig. 5 are computational correlates rather than validated clinical markers. The energy interpretation in Sec. 2.2.1 holds only under symmetric coupling, zero natural frequencies and slowly varying inputs. The hyperparameters K, τ, D and B are calibrated per modality, the delay τ is fixed rather than learned, and the sequential per-step oscillator dynamics make S^2-Net the slowest of the compared models at inference (Table 13). Future work includes relating the latent spikes to physiological measurements and learning the delays.

Acknowledgments and Disclosure of Funding

This work was supported by the National Institutes of Health (AG091653, AG068399, AG084375) and the Foundation of Hope.

Competing Interests

The authors declare no competing interests.

References

Guillaume Bellec, Darjan Salaj, Anand Subramoney, Robert Legenstein, and Wolfgang Maass. Long short-term memory and learning-to-learn in networks of spiking neurons. *Advances in Neural Information Processing Systems*, 31, 2018.

Alberto Benussi, Francesco Di Lorenzo, Valentina Dell'Era, Maura Cosseddu, Antonella Alberici, Salvatore Caratozzolo, Maria Sofia Cotelli, Anna Micheli, Luca Rozzini, Alessandro Depari, et al. Transcranial magnetic stimulation distinguishes Alzheimer disease from frontotemporal dementia. *Neurology*, 89(7):665–672, 2017.

Susan Y Bookheimer, David H Salat, Melissa Terpstra, Beau M Ances, Deanna M Barch, Randy L Buckner, Gregory C Burgess, Sandra W Curtiss, Mirella Diaz-Santos, Jennifer Stine Elam, et al. The lifespan human connectome project in aging: an overview. *Neuroimage*, 185:335–348, 2019.

Michael Breakspear. Dynamic models of large-scale brain activity. *Nature Neuroscience*, 20(3): 340–352, 2017.

Jared C Bronski, Thomas Carty, and Lee DeVille. Configurational stability for the Kuramoto–Sakaguchi model. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 28(10), 2018.

Richard J Burman, Paul JN Brodersen, Joseph V Raimondo, Arjune Sen, and Colin J Akerman. Active cortical networks promote shunting fast synaptic inhibition in vivo. *Neuron*, 111(22): 3531–3540, 2023.

Joana Cabral, Francesca Castaldo, Jakub Vohryzek, Vladimir Litvak, Christian Bick, Renaud Lambiotte, Karl Friston, Morten L Kringelbach, and Gustavo Deco. Metastable oscillatory modes emerge from synchronization in the brain spacetime connectome. *Communications Physics*, 5(1): 184, 2022.

Lin Chen. Topological structure in visual perception. *Science*, 218(4573):699–700, 1982.

Benjamin Cramer, Yannik Stradmann, Johannes Schemmel, and Friedemann Zenke. The Heidelberg spiking data sets for the systematic evaluation of spiking neural networks. *IEEE Transactions on Neural Networks and Learning Systems*, 33(7):2744–2757, 2020.

Marco Cuturi. Sinkhorn distances: Lightspeed computation of optimal transport. *Advances in neural information processing systems*, 26, 2013.

Tingting Dan, Ziquan Wei, Won Hwa Kim, and Guorong Wu. Exploring the enigma of neural dynamics through a scattering-transform mixer landscape for Riemannian manifold. In *International Conference on Machine Learning*, pages 9976–9990. PMLR, 2024.

Tingting Dan, Jiaqi Ding, and Guorong Wu. Explore brain-inspired machine intelligence for connecting dots on graphs through holographic blueprint of oscillatory synchronization. *Nature Communications*, 16(1):9425, 2025.

Maurizio De Pittà and Nicolas Brunel. Multiple forms of working memory emerge from synapse–astrocyte interactions in a neuron–glia network model. *Proceedings of the National Academy of Sciences*, 119(43):e2207912119, 2022.

Gustavo Deco, Viktor Jirsa, Anthony R McIntosh, Olaf Sporns, and Rolf Kötter. Key role of coupling, delay, and noise in resting brain fluctuations. *Proceedings of the National Academy of Sciences*, 106(25):10302–10307, 2009.

Gustavo Deco, Morten L Kringelbach, Viktor K Jirsa, and Petra Ritter. The dynamics of resting fluctuations in the brain: metastability and its dynamical cortical core. *Scientific Reports*, 7(1): 3095, 2017.

Jiaqi Ding, Tingting Dan, Zhixuan Zhou, and Guorong Wu. Let brain rhythm shape machine intelligence for connecting dots on graphs. In *Advances in Neural Information Processing Systems*, volume 38, pages 11806–11828, 2026.

Daniel J Felleman and David C Van Essen. Distributed hierarchical processing in the primate cerebral cortex. *Cerebral cortex (New York, NY: 1991)*, 1(1):1–47, 1991.

Pascal Fries. A mechanism for cognitive dynamics: neuronal communication through neuronal coherence. *Trends in cognitive sciences*, 9(10):474–480, 2005.

Pascal Fries. Rhythms for cognition: communication through coherence. *Neuron*, 88(1):220–235, 2015.

Klaus Greff, Sjoerd Van Steenkiste, and Jürgen Schmidhuber. On the binding problem in artificial neural networks. *arXiv preprint arXiv:2012.05208*, 2020.

Albert Gu and Tri Dao. Mamba: Linear-time sequence modeling with selective state spaces. *arXiv preprint arXiv:2312.00752*, 2023.

Nellippallil Balakrishnan Harikrishnan and Nithin Nagaraj. When noise meets chaos: Stochastic resonance in neurochaos learning. *Neural Networks*, 143:425–435, 2021.

Clifford R Jack Jr, Matt A Bernstein, Nick C Fox, Paul Thompson, Gene Alexander, Danielle Harvey, Bret Borowski, Paula J Britson, Jennifer L. Whitwell, Chadwick Ward, et al. The Alzheimer's Disease Neuroimaging Initiative (ADNI): MRI methods. *Journal of Magnetic Resonance Imaging: An Official Journal of the International Society for Magnetic Resonance in Medicine*, 27(4): 685–691, 2008.

Ping Ju, Qionglin Dai, Hongyan Cheng, and Junzhong Yang. Dynamics in the Sakaguchi–Kuramoto model with two subpopulations. *Physical Review E*, 90(1):012903, 2014.

Yoshiki Kuramoto. Self-entrainment of a population of coupled non-linear oscillators. In *International symposium on mathematical problems in theoretical physics: January 23–29, 1975, kyoto university, kyoto/Japan*, pages 420–422. Springer, 2005.

Quoc V Le, Navdeep Jaitly, and Geoffrey E Hinton. A simple way to initialize recurrent networks of rectified linear units. *arXiv preprint arXiv:1504.00941*, 2015.

Francesco Locatello, Dirk Weissenborn, Thomas Unterthiner, Aravindh Mahendran, Georg Heigold, Jakob Uszkoreit, Alexey Dosovitskiy, and Thomas Kipf. Object-centric learning with slot attention. *Advances in neural information processing systems*, 33:11525–11538, 2020.

Michael London and Michael Häusser. Dendritic computation. *Annual Review of Neuroscience*, 28 (1):503–532, 2005.

Sindy Löwe, Phillip Lippe, Maja Rudolph, and Max Welling. Complex-valued autoencoders for object discovery. *Transactions on Machine Learning Research*, 2022.

Sindy Löwe, Phillip Lippe, Francesco Locatello, and Max Welling. Rotating features for object discovery. *Advances in Neural Information Processing Systems*, 36, 2023.

Wolfgang Maass. Networks of spiking neurons: the third generation of neural network models. *Neural networks*, 10(9):1659–1671, 1997.

Kenneth Marek, Danna Jennings, Shirley Lasch, Andrew Siderowf, Caroline Tanner, Tanya Simuni, Chris Coffey, Karl Kieburtz, Emily Flagg, Sohini Chowdhury, et al. The Parkinson Progression Marker Initiative (PPMI). *Progress in Neurobiology*, 95(4):629–635, 2011.

Matthew M McGregor and Alexandra B Nelson. Circuit mechanisms of Parkinson's disease. *Neuron*, 101(6):1042–1056, 2019.

Karla L Miller, Fidel Alfaro-Almagro, Neal K Bangerter, David L Thomas, Essa Yacoub, Junqian Xu, Andreas J Bartsch, Saad Jbabdi, Stamatios N Sotiropoulos, Jesper LR Andersson, et al. Multimodal population brain imaging in the UK Biobank prospective epidemiological study. *Nature Neuroscience*, 19(11):1523–1536, 2016.

Takeru Miyato, Sindy Löwe, Andreas Geiger, and Max Welling. Artificial Kuramoto oscillatory neurons. In *The Thirteenth International Conference on Learning Representations*, 2025.

Lyle Muller, Frédéric Chavane, John Reynolds, and Terrence J Sejnowski. Cortical travelling waves: mechanisms and computational principles. *Nature Reviews Neuroscience*, 19(5):255–268, 2018.

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

Tuan Nguyen, Hirotada Honda, Takashi Sano, Vinh Nguyen, Shugo Nakamura, and Tan Minh Nguyen. From coupled oscillators to graph neural networks: Reducing over-smoothing via a Kuramoto model-based approach. In *International Conference on Artificial Intelligence and Statistics*, pages 2710–2718. PMLR, 2024.

Jorge J Palop and Lennart Mucke. Network abnormalities and interneuron dysfunction in Alzheimer disease. *Nature Reviews Neuroscience*, 17(12):777–792, 2016.

Vincent Planche, Boris Mansencal, José V Manjon, Thomas Tourdias, Gwenaëlle Catheline, Pierrick Coupé, Frontotemporal Lobar Degeneration Neuroimaging Initiative, and the National Alzheimer's Coordinating Center cohort. Anatomical MRI staging of frontotemporal dementia variants. *Alzheimer's & Dementia*, 19(8):3283–3294, 2023.

Rajesh PN Rao and Dana H Ballard. Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nature neuroscience*, 2(1):79–87, 1999.

Pieter R Roelfsema, Victor AF Lamme, and Henk Spekreijse. Object-based attention in the primary visual cortex of the macaque monkey. *Nature*, 395(6700):376–381, 1998.

Pieter R Roelfsema. Solving the binding problem: Assemblies form when neurons enhance their firing rate—they don't need to oscillate or synchronize. *Neuron*, 111(7):1003–1019, 2023.

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

T Konstantin Rusch, Ben Chamberlain, James Rowbottom, Siddhartha Mishra, and Michael Bronstein. Graph-coupled oscillator networks. In *International conference on machine learning*, pages 18888–18909. PMLR, 2022.

Maximilian Seitzer, Max Horn, Andrii Zadaianchuk, Dominik Zietlow, Tianjun Xiao, Carl-Johann Simon-Gabriel, Tong He, Zheng Zhang, Bernhard Schölkopf, Thomas Brox, and Francesco Locatello. Bridging the gap to real-world object-centric learning. In *The Eleventh International Conference on Learning Representations*, 2023.

Wolf Singer. Neuronal synchrony: a versatile code for the definition of relations? *Neuron*, 24(1): 49–65, 1999.

Aleksandar Stanić, Anand Gopalakrishnan, Kazuki Irie, and Jürgen Schmidhuber. Contrastive training of complex-valued autoencoders for object discovery. *Advances in Neural Information Processing Systems*, 36, 2023.

Aida Todri-Sanial, Corentin Delacour, Madeleine Abernot, and Filip Sabo. Computing with oscillators from theoretical underpinnings to applications and demonstrators. *Npj unconventional computing*, 1(1):14, 2024.

Nathalie Tzourio-Mazoyer, Brigitte Landeau, Dimitri Papathanassiou, Fabrice Crivello, Octave Etard, Nicolas Delcroix, Bernard Mazoyer, and Marc Joliot. Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *Neuroimage*, 15(1):273–289, 2002.

David C Van Essen, Stephen M Smith, Deanna M Barch, Timothy EJ Behrens, Essa Yacoub, Kamil Ugurbil, WU-Minn HCP Consortium, et al. The WU-Minn Human Connectome Project: an overview. *Neuroimage*, 80:62–79, 2013.

Francisco Varela, Jean-Philippe Lachaux, Eugenio Rodriguez, and Jacques Martinerie. The brainweb: phase synchronization and large-scale integration. *Nature reviews neuroscience*, 2(4):229–239, 2001.

Pete Warden. Speech commands: A dataset for limited-vocabulary speech recognition. *arXiv preprint arXiv:1804.03209*, 2018.

Yujie Wu, Lei Deng, Guoqi Li, Jun Zhu, Yuan Xie, and Luping Shi. Direct training for spiking neural networks: Faster, larger, better. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 1311–1318, 2019.

Yinsong Yan, Qu Yang, Yujie Wu, Hanwen Liu, Malu Zhang, Haizhou Li, Kay Chen Tan, and Jibin Wu. Efficient and robust temporal processing with neural oscillations modulated spiking neural networks. *Nature Communications*, 16(1):8651, 2025.

MK Stephen Yeung and Steven H Strogatz. Time delay in the Kuramoto model of coupled oscillators. *Physical review letters*, 82(3):648, 1999.

Bojian Yin, Federico Corradi, and Sander M Bohté. Effective and efficient computation with multiple-timescale spiking recurrent neural networks. In *International Conference on Neuromorphic Systems 2020*, pages 1–8, 2020.

Bojian Yin, Federico Corradi, and Sander M Bohté. Accurate and efficient time-domain classification with adaptive spiking recurrent neural networks. *Nature Machine Intelligence*, 3(10):905–913, 2021.

Friedemann Zenke and Tim P Vogels. The remarkable robustness of surrogate gradient learning for instilling complex function in spiking neural networks. *Neural Computation*, 33(4):899–925, 2021.

Hao Zheng, Hui Lin, Rong Zhao, and Luping Shi. Dance of SNN and ANN: Solving binding problem by combining spike timing and reconstructive attention. *Advances in Neural Information Processing Systems*, 35, 2022.

Hao Zheng, Hui Lin, and Rong Zhao. GUST: Combinatorial generalization by unsupervised grouping with neuronal coherence. *Advances in Neural Information Processing Systems*, 36, 2023.

A Appendix

A.1 Dataset Descriptions

In this section, we describe the datasets used to evaluate the proposed model across brain state recognition, disease diagnosis, temporal, auditory, neuromorphic, and sequence modeling tasks.

1. The Lifespan Human Connectome Project Aging (HCP-A) dataset (Bookheimer et al., 2019). The HCP-A dataset characterizes brain functional and structural organization across the aging process and is widely used for task recognition studies. It consists of data from 717 subjects, including both task-based fMRI scans (4,846 time series in total) and diffusion-weighted imaging (DWI) scans (717 in total). The dataset includes four task conditions related to memory and sensorimotor processing: VISMOTOR, CARIT, FACENAME, and Resting State. Each fMRI scan contains 300 time points. In our experiments, these task conditions are treated as distinct categories in a four-class classification problem.

2. The Human Connectome Project – Young Adults (HCP-YA) dataset (Van Essen et al., 2013). The HCP-YA dataset contains high-quality task-based and resting-state fMRI data from 1,200 healthy young adults (each includes two scans). It includes seven cognitive tasks: Motor, Relational, Social, Working Memory, Language, Emotion, and Gambling. Each task-based fMRI scan consists of 175 time points. In addition, the Working Memory task (denoted as HCP-WM) alternates between 2-back and 0-back conditions with body, place, face, and tool stimuli, interleaved with fixation periods. A resting-state period follows every two task blocks, resulting in fMRI scans with 405 time points. Data preprocessing follows the pipeline described in (Dan et al., 2024). For both HCP-A and HCP-YA datasets, the brain is parcellated into 116 regions using the AAL atlas (Tzourio-Mazoyer et al., 2002). Structural connectivity (SC) is represented as a 116×116 matrix, where each entry denotes the number of reconstructed white-matter fibers between two regions, normalized by the total fiber count of each subject. Functional connectivity (FC) is computed using Pearson correlation between regional BOLD time series.

In our experiments, the HCP-YA tasks are treated as a seven-class classification problem, while the HCP-WM conditions form a nine-class classification problem. We report results using 10-fold cross-validation.

3. UK Biobank (UKB) dataset (Miller et al., 2016) (application number 91673). The UK Biobank is a large-scale population-based cohort that provides multimodal neuroimaging data from 16,240 participants. We use task-based fMRI data from the UKB imaging subset for task recognition. The dataset includes multiple cognitive paradigms designed to probe visual, motor, and higher-order cognitive functions. Each scan contains temporally aligned BOLD time series acquired under standardized imaging protocols. Similar to the HCP datasets, FC is computed from regional BOLD signals, and SC is derived when diffusion data are available. The UKB tasks are treated as distinct classes in a multi-class classification setting, enabling evaluation of model generalization on a large and heterogeneous population.

4. Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset (Jack Jr et al., 2008). The ADNI dataset is a longitudinal, multi-site study aimed at understanding the progression of Alzheimer's disease (AD). We use resting-state fMRI and diffusion MRI data from 135 subjects spanning cognitively normal (CN), mild cognitive impairment (MCI), and Alzheimer's disease (AD) groups. Functional connectivity is constructed from resting-state BOLD signals, while structural connectivity is estimated from diffusion tractography. The diagnostic labels are used to formulate a disease classification task, assessing the ability of the model to distinguish neurodegenerative stages.

5. Parkinson's Progression Markers Initiative (PPMI) dataset (Marek et al., 2011). The PPMI dataset focuses on Parkinson's disease (PD) and related disorders, providing multimodal imaging data including resting-state fMRI and diffusion MRI. This multi-center study provides neuroimaging data from 173 subjects, including individuals with Parkinson's disease (PD), scans without evidence of dopaminergic deficit (SWEDD), prodromal PD, and CN. Similar preprocessing and connectivity construction procedures are applied as in ADNI. The resulting FC and SC representations are used for four-class diagnosis (PD, SWEDD, prodromal and CN), evaluating the sensitivity of the proposed method to Parkinson's-related network alterations.

6. Neuroimaging in Frontotemporal Dementia (NIFD) dataset (Planche et al., 2023). This dataset focuses on frontotemporal dementia (FTD) and includes resting-state fMRI data from 1,010

subjects. Participants are categorized into CN, logopenic variant of primary progressive aphasia (LVPPA), behavioral variant frontotemporal dementia (BV), progressive non-fluent aphasia (PNFA), and semantic variant (SV) groups. We use resting-state fMRI data to construct functional brain networks, capturing disease-related disruptions in large-scale connectivity patterns. The task is formulated as five-class classification (CN and the four FTD variants), providing an additional benchmark for evaluating model robustness across different neurodegenerative diseases.

7. S-MNIST and PS-MNIST (Le et al., 2015). Both datasets are derived from the original MNIST digit recognition benchmark by converting each 28×28 image into a temporal sequence of length 784 via raster scanning. For S-MNIST, pixels are fed in natural scan order, while PS-MNIST applies a fixed random permutation to remove spatial locality. Pixel intensities are injected as input current to the encoding layer, transforming non-spiking inputs into spike-based sequences for downstream SNN processing.

8. SHD (Cramer et al., 2020). The Heidelberg Digits dataset contains $\sim$10,000 audio recordings of spoken digits (0–9) in English and German, captured from 12 speakers. Each speaker utters each digit $\sim$40 times across both languages. Audio is transformed into spike trains using a biologically inspired cochlea model. Following standard practice, each sequence is represented as 1000 time steps. The dataset is split into 8156 training and 2264 test samples.

9. ECG (Yin et al., 2021). This dataset consists of six types of ECG waveforms (P, PQ, QR, RS, ST, TP). We follow the preprocessing procedure in prior work, applying a variant of the level-crossing encoding method to the derivative of the normalized ECG signal. Each channel is converted into two spike trains capturing value-increasing and value-decreasing events, respectively.

10. GSC (Warden, 2018). The Google Speech Commands dataset (GSC) contains 64,727 utterances from 1881 speakers. Following standard configuration, we use 12 classes selected from the 35 available labels: "Yes", "No", "Up", "Down", "Left", "Right", "On", "Off", "Stop", "Go", along with "Unknown" and "Silence". Audio preprocessing follows prior work, extracting log Mel filterbank features and derivatives. We use GSC as a keyword-spotting benchmark for speech recognition.

A.2 Energy Interpretation and Lyapunov Stability Analysis

To interpret the graph-aware vector Kuramoto dynamics, we describe them in terms of a potential energy. We analyze Eq. (2) with $\omega = 0$ and show that its interaction and forcing terms are the negative gradient of a potential $V(\theta, t)$. With nonzero, heterogeneous natural frequencies the system is no longer a gradient flow, and the result below does not apply directly.

Theorem 1. *Under the assumptions of symmetric structural coupling ($A_{ij} = A_{ji}$), antisymmetric phase lags ($\alpha_{ij} = -\alpha_{ji}$) and $\omega = 0$, the interaction and sensory drive components of the update rule follow the negative gradient of the following scalar potential field:*

$$\begin{aligned} V(\theta, t) = & \underbrace{-\frac{K}{2N} \sum_{d=1}^{D} \sum_{i=1}^{N} \sum_{j=1}^{N} A_{ij} \cos\big(\theta_{j,d} - \theta_{i,d} - \alpha_{ij}\big)}_{V_{\text{sync}}\text{: System dynamics}} \\ & \underbrace{- \sum_{d=1}^{D} \sum_{i=1}^{N} \kappa_{i,d} \cos\big(\gamma_i(t) - \theta_{i,d}\big)}_{V_{\text{data}}\text{: System control}} . \end{aligned} \tag{8}$$

Proof. Taking the partial derivative of V with respect to the d-th phase component of node i, denoted $\theta_{i,d}$, we obtain:

$$\begin{aligned} -\frac{\partial V}{\partial \theta_{i,d}} = & \frac{\partial}{\partial \theta_{i,d}} \left[\sum_{m=1}^{N} \kappa_{m,d} \cos(\gamma_m - \theta_{m,d}) \right] \\ & + \frac{K}{2N} \frac{\partial}{\partial \theta_{i,d}} \left[\sum_{m,n} A_{mn} \cos(\theta_{n,d} - \theta_{m,d} - \alpha_{mn}) \right] . \end{aligned} \tag{9}$$

Leveraging the symmetry of A and antisymmetry of α, the interaction derivative simplifies, yielding:

$$-\frac{\partial V}{\partial \theta_{i,d}} = \kappa_{i,d}\sin(\gamma_i(t) - \theta_{i,d}) + \frac{K}{N}\sum_{j=1}^{N} A_{ij}\sin(\theta_{j,d} - \theta_{i,d} - \alpha_{ij}). \tag{10}$$

This result exactly matches the update terms in our dynamic equation (Eq. 2), confirming that, for $\omega = 0$, the trajectory satisfies $\dot{\theta}_{i,d} = -\partial V/\partial\theta_{i,d}$. Treating γ as frozen, i.e., neglecting $\partial V/\partial t$, which is justified when it varies slowly relative to the relaxation time of the oscillator network, the energy is non-increasing along the trajectory:

$$\frac{dV}{dt} = \sum_{i,d}\frac{\partial V}{\partial \theta_{i,d}}\dot{\theta}_{i,d} = -\sum_{i,d}\left(\frac{\partial V}{\partial \theta_{i,d}}\right)^2 \leq 0, \tag{11}$$

so that, under these assumptions, the phases converge to a critical point of V. □

Physical interpretation. The potential V balances two objectives:

Internal Coherence (V_{sync}): The first term forms a "topological prior." It is minimized when the phase difference $\Delta\theta_{ij}$ between connected regions matches the lag α_{ij}, whose magnitude is set by the structural connectivity and whose sign is learned (Eq. 7). This term therefore biases the latent phases toward configurations that are consistent with the structural connectivity.

Data Consistency (V_{data}): The second term acts as a "data anchor." Modeled as a sinusoidal potential well centered at $\gamma_i(t)$, it pulls the oscillator state toward the current sensory observation. The stiffness parameter $\kappa_{i,d}$ determines the confidence in the observation.

Under the assumptions above, minimizing V trades off consistency with the structural prior against agreement with the current input.

Remark: learnable topology. For general temporal datasets without a predefined graph, we treat the adjacency matrix $A \in \mathbb{R}^{N\times N}$, which governs the coupling strength between N hidden neurons, as a learnable parameter initialized from a random distribution. We keep A symmetric ($A_{ij} = A_{ji}$) so that the energy interpretation above still applies when $\omega = 0$. To that end, we parameterize a weight matrix W_A and compute $A = \sigma(W_A + W_A^\top)/2$, where σ denotes the sigmoid function. The direction of propagation is therefore not encoded in A but in the antisymmetric phase lag α_{ij} (see below), which allows directed, traveling-wave-like coordination. Through end-to-end training, connections that help the task are strengthened and the others are weakened, so that S^2-Net learns a topology adapted to the input domain:

- For **visual tasks (e.g., S-MNIST, PS-MNIST)**, we expect the learned matrix to capture the latent *spatial geometry*, mapping implicit temporal offsets (e.g., vertically adjacent pixels) into explicit couplings.
- For **audio/speech tasks (e.g., SHD, GSC)**, we expect the learned couplings to reflect which frequency bands tend to co-activate (e.g., the harmonic structure of vowels), facilitating the integration of distributed acoustic cues into words.

Our hypothesized interpretation for each benchmark is summarized in Table 3.

The phase lag is computed as for the brain datasets (Eq. (7)), with the learned A in place of the structural connectivity: $\alpha_{ij} = \alpha_0\tanh(W_{ij} - W_{ji})\,\tilde{C}_{ij}$, where $\tilde{C}_{ij}$ is the min–max normalized transport cost $C_{ij} \propto 1/(A_{ij}+\varepsilon)$. The antisymmetry $\alpha_{ij} = -\alpha_{ji}$ required by the energy interpretation thus holds by construction, and $\tanh$ keeps the lag bounded.

A.3 Deploy S^2-Net in Downstream Applications

We utilize three distinct heads to drive feature representation learning.

(1) Self-Supervised Reconstruction. To preserve intrinsic information within the hidden representations, a decoder f_{dec} acts as a regularizer by reconstructing the original input from the latent membrane-like state $U(t)$:

$$\hat{X}(t) = f_{\text{dec}}(U(t)), \quad \mathcal{L}_{\text{rec}} = \sum_t \|X(t) - \hat{X}(t)\|_2^2 + \lambda\sum_t \|\hat{X}(t) - \hat{X}(t-1)\|_2^2 \tag{12}$$

Table 3: Hypothesized interpretation of the learned adjacency matrix A and the role of synchronization across benchmarks (not verified quantitatively).

Dataset	**Input Domain**	**Physical Meaning of A**	**Function (Role of Synchronization)**
S/PS-MNIST	Visual (Flattened)	Spatial Geometry	Reconstructs latent 2D spatial shapes from flattened 1D temporal sequences.
SHD	Auditory (Spikes)	Tonotopic Binding	Integrates harmonics across frequency bands to form coherent auditory objects.
ECG	Physiological (1D)	Phase/Morphology	Tracks the cyclic phase transitions of heartbeats and captures waveform morphology.
GSC	Speech (Spectrogram)	Spectro-temporal co-activation	Groups the acoustic cues of a spoken keyword.

where λ controls the strength of the smoothness regularization.

(2) Dynamic State Recognition. For the sequence labeling task, we require a prediction for state status $\hat{y}^{(\text{state})}(t)$ for every time step t. The accumulated membrane potential $\mathcal{H}(t)$ of the readout layer is mapped to class probabilities via the projection f^1_{cls}:

$$\hat{y}^{(\text{state})}(t) = f^1_{\text{cls}}(\mathcal{H}(t)), \quad \hat{y}^{(\text{state})}(t) \in \mathbb{R}^C, \tag{13}$$

where C is the number of dynamic states.

(3) Classification. For subject-level classification, we aggregate temporal information by pooling the accumulated logits $\mathcal{H}(t)$ across the sequence length T and predict the label $\hat{y}^{(\text{cls})}$ through project f^2_{cls}:

$$\hat{y}^{(\text{cls})} = f^2_{\text{cls}}\left(\frac{1}{T}\sum_{t=1}^{T}\mathcal{H}(t)\right). \tag{14}$$

A.4 Additional Experimental Results

A.4.1 Simulated Experiments

Experimental setup. To validate *binding-by-synchrony* and *traveling-wave propagation*, we constructed two synthetic environments (Fig. 6): (a) disjoint squares and (b) nested topological structures[3]. Objects are initialized with distinct intensities (1.0 vs. 0.5) to enable feature-driven repulsion via the graph coupling term A_{ij}. We applied Gaussian perturbation at varying levels: a baseline $\sigma = 0.05$ for (a), and a severe noise regime $\sigma = 0.2$ for (b) to test robustness under low signal-to-noise ratio. Each of the $N = 28 \times 28 = 784$ pixels is one unit. In the spike rasters, the vertical axis is the flattened pixel index and the horizontal axis is the simulation step. Unlike S-MNIST, the image is not scanned as a sequence.

Evaluation metrics. We quantify binding quality and separation efficacy using Kuramoto-derived metrics (i.e., Kuramoto order parameter, KOP). Let $\mathcal{O}_q$ be the set of M_q pixels for object q. *Intra-Object Synchronization (R_{intra})* measures internal coherence: $R_{intra} = \frac{1}{Q}\sum_{q=1}^{Q}\left|\frac{1}{M_q}\sum_{j\in\mathcal{O}_q} e^{i\theta_j}\right|$.

[3]In computational neuroscience, nested structures are regarded as a canonical benchmark (Roelfsema et al., 1998; Chen, 1982) for figure-ground segregation. They challenge the system to solve the "binding problem" by integrating spatially separated segments into a coherent whole.

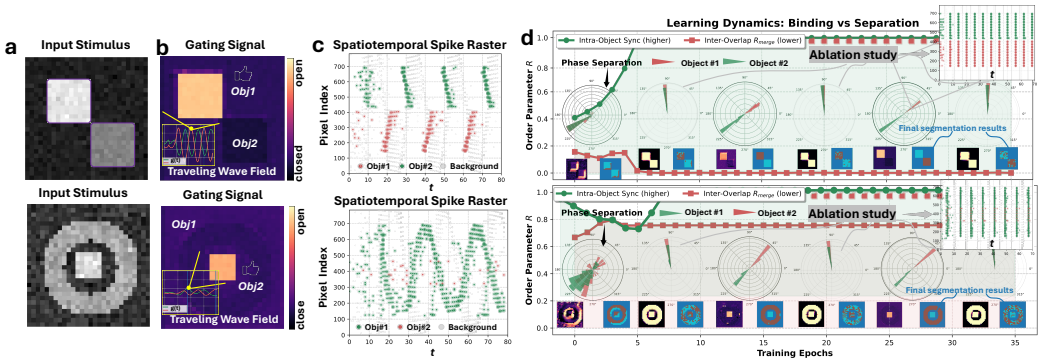

Figure 6: Proof-of-concept simulations of S^2-Net. (a) The input stimulus: disjoint object discovery (top) and nested topological binding (bottom). (b) The gating signal shows Object #1 or (#2) in an "Open" phase (bright) while Object #2 or (#1) is "Closed" (dark). Corresponding gating dynamics manifest as alternating sine waves, resembling rhythmic attentional sampling. (c) Spatiotemporal spike raster: graph-structured delays produce a traveling-wave field, and the two objects keep distinct temporal representations despite spatial entanglement. (d) Evolutionary curves of the order parameter R quantify the learning process. Insets visualize the transition from initial phase chaos to stable separation (polar plots) and the final segmentation results.

A value near 1.0 indicates perfect binding. *Inter-Object Overlap (R_{merge})* assesses desynchronization by treating the union of objects as a single population: $R_{merge} = \left| \frac{1}{\sum_{q=1}^{Q} M_q} \sum_{j \in \bigcup \mathcal{O}_q} e^{i\theta_j} \right|$. Ideally, $R_{merge} \to 0$ represents successful phase separation. Since the simulations use $D = 1$, each θ_j is a scalar angle. Note that R_{merge} is also low for random phases and is nonzero for perfectly separated objects of unequal size (for two anti-phase objects, $R_{merge} = |M_1 - M_2|/(M_1 + M_2)$), so it is not a stand-alone measure of separation.

Discrete object discovery and attention rhythms (Fig. 6-top) In the disjoint task (two objects shown in Fig. 6a), our S^2-Net breaks phase symmetry, driving the system from a disordered transient state into a stable anti-phase regime. The gating dynamics $g(t)$ (Fig. 6b) emerge as smooth, alternating sinusoidal waves. This creates a "push-pull" mechanism that resembles rhythmic attentional sampling. The resulting spatiotemporal spike raster (Fig. 6c) reveals temporal chunking, where objects are phase-locked to alternating firing windows separated by refractory gaps. As shown in Fig. 6d, R_{intra} rapidly converges to ≈ 1.0 while R_{merge} collapses to ≈ 0.0, consistent with segregation of the two objects.

Topological complexity and balanced competition (Fig. 6-bottom). Standard coupled oscillators often collapse into in-phase synchronization or a winner-take-all regime where the larger ring suppresses the smaller square (Fig. 6a). In contrast, S^2-Net keeps both objects represented through graph-aware traveling waves. As shown in the gating signal map (Fig. 6b), a phase gradient propagates along the ring, binding its spatially disjoint segments into a unified entity. As a result, the smaller square is not suppressed despite the size imbalance and competes equally for representation. The spike raster (Fig. 6c) reveals a traveling wave topology, where the square fires precisely within the local phase troughs of the ring's propagation, resolving the topological conflict through autonomous temporal multiplexing.

Ablation and dynamical stability. Upon turning off the graph-aware coupling and phase lag, the system immediately undergoes global synchronization collapse, with both R_{intra} and R_{merge} converging to 1.0. This indicates that phase separation depends on the graph-aware coupling and the phase lag.

Multi-object benchmarks and metrics. To complement the two-object simulations, we generate 32×32 single-channel grayscale scenes with K objects drawn from {square, ellipse, triangle}, each with random scale, orientation $\phi \sim \mathcal{U}(0, \pi)$, position and intensity $\sim \mathcal{U}(0.5, 1.0)$ on a zero background. In the overlapping dSprites-style variant, objects may occlude one another, the visible foreground mask defines the target segmentation, and Gaussian noise with $\sigma = 0.1$ is added; we also use a non-overlapping, grid-placed variant (*shapes*). We report FG-ARI on foreground pixels with K clusters, full-scene ARI on all pixels with $K + 1$ clusters (the background as a separate segment),

and within-object coherence, i.e., the mean resultant length of the phases inside each object, minus the value obtained with random phases under the same masks. Unlike R_{merge}, these metrics do not depend on object size and are close to zero for random phases.

Table 4: Phase-lag ablation and comparison with AKOrN on dSprites-style scenes ($K = 3$, $\sigma = 0.1$, 3 seeds).

Model	FG-ARI	Full-scene ARI
S^2-Net (full)	0.28	**0.54**
S^2-Net without phase lag ($\alpha = 0$)	0.13	0.52
AKOrN (Miyato et al., 2025)	0.30	0.26
AKOrN + our phase lag and delay ($\tau = 2$)	**0.34**	0.33

Results. Removing the phase lag halves FG-ARI, while full-scene ARI barely changes, so the lag mainly separates the objects from each other (Table 4). AKOrN reaches a similar FG-ARI but a much lower full-scene ARI, because it does not assign the background a coherent phase; adding our phase lag and delay to AKOrN improves both scores, which suggests that this inductive bias transfers to other oscillator models. When the model is trained with $K = 3$ and evaluated with $K = 2, \ldots, 6$, both ARIs decrease as K grows (e.g., full-scene ARI from 0.57 to 0.44 on dSprites and from 0.60 to 0.38 on shapes), whereas chance-corrected coherence stays stable (0.31–0.33 for dSprites and 0.51–0.52 for shapes). The raw coherence rises slightly with K only because the random-phase floor grows from 0.12 to 0.16 as objects become smaller. Starting from random phases on a 4-object scene, within-object coherence and FG-ARI rise over the Kuramoto iterations and saturate after about 24 steps (coherence $0.12 \to 0.98$, FG-ARI $0.00 \to 0.33$), so grouping emerges progressively rather than by instantaneous locking. Untrained couplings collapse into global synchrony and random phases give FG-ARI ≈ 0. Trained models avoid global locking (global order parameter 0.20–0.35) while keeping high within-object (> 0.80) and low between-object (< 0.15) synchrony, i.e., the phases organize at the scale of objects rather than as a single global traveling wave.

A.4.2 Extended Baselines Experiments

To evaluate S^2-Net beyond the neuromorphic domain, we also compare it with general-purpose and physics-inspired deep learning models.

Specifically, we consider three categories of models:

- **Standard Deep Sequence Models:** Transformer, LSTM, and Temporal Convolutional Networks (TCN).
- **State-Space Models:** Mamba, a selective state-space model for long sequences.
- **Physics-Inspired and Synchronization Baselines:** AKOrN, GraphCON, KuramotoGNN, and stochastic resonance (SR)-Net, which are continuous-time or oscillator-based neural networks.

For a fair comparison, all baseline models use the same settings: a hidden dimension of 64, a batch size of 16, and a maximum of 500 training epochs. We report the 10-fold cross-validation performance measured by Accuracy, Precision, and F1-score across six diverse neuroimaging datasets (HCPA, HCPYA, UKB, ADNI, PPMI, and NIFD).

As shown in Table 5, S^2-Net outperforms both the sequence models and the synchronization-based networks on most datasets. Standard sequence models are competitive on the large task-fMRI datasets but drop markedly on the small ADNI cohort, where S^2-Net remains accurate. On PPMI, LSTM and Mamba reach a higher F1 than S^2-Net (44.82% and 44.86% vs. 41.51%), so the advantage there is limited to accuracy and precision.

A.5 Ablation Study

We ablate the components of S^2-Net (Table 6), test whether spiking matters beyond capacity (Table 7), compare reset schemes (Table 8), vary the number of dendritic branches (Table 9), and re-evaluate the SNN baselines under a capacity-matched protocol (Table 10).

Table 5: Performance of nine models. 10-fold cross-validation results are reported for Accuracy, Precision, and F1-score (%) across six brain-network datasets.

Model	Metric	HCPA	HCPYA	UKB	ADNI	PPMI	NIFD
Transformer	Acc	96.89±0.39	96.48±0.22	98.83±0.26	78.17±3.76	66.10±2.68	61.11±2.03
	Pre	96.14±0.67	96.46±0.27	96.87±0.26	71.72±18.89	45.01±5.16	57.45±4.21
	F1	95.85±0.63	96.47±0.24	96.81±0.26	64.44±12.65	41.79±1.76	54.17±3.58
LSTM	Acc	97.26±0.70	95.39±0.24	98.95±0.24	74.78±6.62	66.15±4.72	62.74±2.30
	Pre	96.27±1.34	95.27±0.24	97.03±0.25	67.72±14.87	45.38±3.72	58.51±3.42
	F1	96.27±0.90	95.28±0.24	96.94±0.24	63.62±12.31	44.82±3.34	55.38±3.63
TCN	Acc	97.84±0.48	96.65±0.23	96.01±0.23	76.07±5.78	67.22±5.28	63.12±3.39
	Pre	97.21±0.60	96.62±0.28	96.08±0.22	63.56±3.89	40.63±12.19	62.39±1.87
	F1	97.12±0.67	96.62±0.27	96.00±0.22	59.22±15.25	40.75±9.33	54.35±3.77
Mamba	Acc	98.27±0.10	96.19±0.14	96.97±0.24	70.54±6.58	67.49±1.69	63.46±2.94
	Pre	98.42±0.03	96.15±0.16	97.03±0.23	52.98±19.41	47.46±3.51	60.51±4.04
	F1	97.72±0.20	96.13±0.14	96.96±0.23	51.77±13.06	44.86±1.09	56.38±4.10
AKOrN	Acc	94.34±0.76	86.31±1.71	92.22±1.68	85.27±7.52	66.82±4.87	58.79±2.68
	Pre	94.41±0.85	86.61±1.81	93.15±1.20	70.24±11.73	46.86±10.79	58.74±3.16
	F1	94.30±0.81	86.23±1.73	91.88±1.65	76.46±10.28	39.81±8.82	58.31±3.13
GraphCON	Acc	93.72±0.59	73.93±2.71	56.89±1.16	84.89±5.03	61.34±2.31	49.97±1.25
	Pre	93.61±0.56	74.13±2.41	57.35±1.53	74.46±10.68	47.06±8.98	33.26±7.36
	F1	93.53±0.57	73.67±2.51	56.26±1.41	73.34±8.13	41.19±5.04	37.89±4.32
KuramotoGNN	Acc	86.81±1.35	60.23±3.77	37.23±4.97	72.97±6.45	57.75±4.09	47.43±0.79
	Pre	87.56±0.79	61.97±2.68	40.09±1.47	73.32±6.53	47.89±13.21	44.14±2.24
	F1	85.47±1.33	59.97±3.46	36.23±4.46	73.90±5.83	41.21±9.18	43.87±2.79
SR-Net	Acc	60.41±2.12	36.79±3.18	59.24±1.40	71.81±3.27	57.41±2.50	29.37±2.63
	Pre	45.45±2.85	36.45±3.14	58.60±1.57	41.00±15.90	38.42±11.65	23.79±3.66
	F1	47.25±2.35	35.63±3.16	58.57±1.54	43.17±4.98	35.88±8.39	19.13±1.88

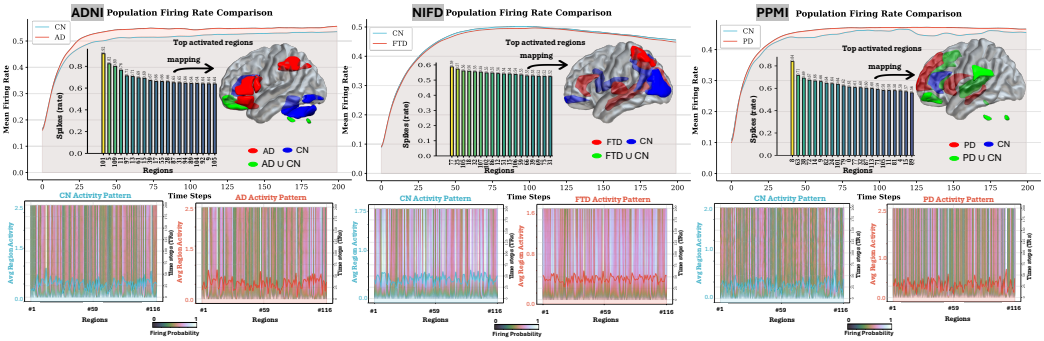

Figure 7: Latent spiking representations and the brain regions that contribute most to classification. The figure is organized into three dataset-specific columns: ADNI, NIFD, and PPMI, arranged from global dynamics to regional patterns: Top Row. Mean firing rate curves for healthy control (CN, blue) and disease (red) groups. Middle Rows. Ranked brain region bar charts based on classification contribution (quantified by spiking intensity (rates)) and their corresponding anatomical mapping to 3D brain projections. Bottom Row. Spatiotemporal raster plots (brain regions vs. time steps) visualizing local spiking event density (firing probability from 0 to 1).

Network components. Table 6 adds the components of S^2-Net one at a time on HCP-YA, starting from the dendritic SNN backbone. Adding Kuramoto oscillators raises F1 from 89.5% to 91.3%, while the phase lag alone changes little (91.4%) before the structural prior is introduced. The largest step, and the only significant one, comes from the structural connectivity (F1 95.2%, $p < 0.05$): with A, the antisymmetric lags of Eq. (7) organize phase differences along connected regions, consistent with the energy interpretation in Eq. (3). Rhythmic gating adds a further 0.3 points, and the time delay τ, which converts phase relations into a temporal order of gating windows, gives the full model (F1 97.0%).

Spiking core versus a non-spiking recurrent core. Simply removing the spiking layer is not an adequate control, so we replace only the spiking core with a threshold-linear ReLU-RNN of matched parameter count (Table 7). The spiking core is 3.25 points more accurate and its mean activity is 41%

Table 6: Component ablation on HCP-YA (%). Each row adds one component to the row above; *: significant improvement over the previous configuration ($p < 0.05$).

Configuration	Acc	Pre	F1
SNN backbone	91.9	92.3	89.5
+ Kuramoto oscillators	92.9	93.7	91.3
+ phase lag	92.9	93.8	91.4
+ structural connectivity*	95.0	95.2	95.2
+ rhythmic gating	95.2	95.5	95.5
+ time delay (full S^2-Net)	**97.0**	**97.1**	**97.0**

Table 7: Spiking core versus a capacity-matched ReLU-RNN core on PPMI. Both receive the same gated phase features; folds, optimizer and training budget are identical.

Temporal core	Params	Acc (%)	Mean activity	Time / fold (min)
SNN (S^2-Net)	156,451	**67.62±5.21**	**0.448**	32.5
ReLU-RNN	156,216	64.37±5.43	0.758	29.0

lower (0.448 vs. 0.758), at a modest increase in training time. The ReLU-RNN does not fail, which indicates that the oscillatory front end carries much of the task-relevant information; the spiking core adds accuracy and sparse, event-based activity, and it produces the latent spike patterns analyzed in Fig. 5.

Table 8: Reset scheme in a rhythm-gated LIF population (5 seeds). Run length: mean number of consecutive spikes; off-cycle rate: firing outside the gating windows.

Reset scheme	Run length	Firing rate	Off-cycle rate
Reset inside the gate	4.27±0.01	0.658	0.658
Reset outside the gate (ours, Eq. (6))	1.01±0.00	0.298	0.501
Hard reset	1.00±0.00	0.261	0.502
Soft reset	1.64±0.00	0.546	0.503

Reset scheme. We compare where the post-spike reset is applied in a rhythm-gated LIF population (5 seeds), measuring the mean number of consecutive spikes (Table 8). If the reset is applied inside the gate, a neuron that has fired keeps its supra-threshold potential while the gate is closed and fires repeatedly, producing runs of 4.27 spikes. Applying the reset outside the gate, as in Eq. (6), removes these runs (1.01), similar to a hard reset (1.00) and shorter than a soft reset (1.64).

Number of dendritic branches. A single leaky branch has a monotonically decaying impulse response, whereas several branches with different decay rates can form non-monotone temporal kernels. Going from one to two branches gives the largest gain on both datasets (accuracy +3.7 points on PPMI and +1.3 points on NIFD; Table 9). With more than two branches, the gains are within one standard deviation, while the parameter count grows from 156K ($B = 4$) to 265K ($B = 8$). We therefore use $B = 4$ as a trade-off between capacity and parameter efficiency.

Capacity-matched comparison. The SNN baselines in Table 1 use the default settings of Yan et al. (2025), including a hidden dimension of 1024 (Table 12), whereas S^2-Net uses 64. To separate the effect of the architecture from capacity and tuning, we retrain all SNN baselines on PPMI with the setting of S^2-Net (Table 10). Under this matched budget, most baselines lose accuracy (e.g., DH-SNN from 65.40% to 61.85% and Rhythm-SNN from 65.57% to 64.21%), and S^2-Net has the highest accuracy, precision and F1, with LSNN closest in accuracy (67.26%). Conversely, increasing the hidden dimension of S^2-Net to 1024 raises its accuracy to 68.49±3.31%, so its advantage does not come from additional capacity.

Sensitivity Analysis of Hyperparameters. To evaluate the robustness of S^2-Net, we conducted a sensitivity analysis on the PPMI dataset, focusing on three components of the oscillatory dynamics: coupling strength (K), time delay (τ), and oscillation dimension (D). As reported in Table 11, S^2-Net is robust to moderate changes, and the default configuration ($K = 1.0$, $\tau = 2$, $D = 4$) performs well overall.

Table 9: Effect of the number of dendritic branches B (%).

Dataset	B	Params	Acc	Pre	F1
PPMI	1	75K	64.15±4.14	46.28±6.45	38.92±5.85
	2	102K	67.85±4.85	49.65±7.12	41.78±6.45
	4	156K	67.62±5.21	49.30±8.83	41.51±7.19
	8	265K	68.35±3.80	50.25±6.05	42.45±5.15
NIFD	1	75K	61.98±2.01	59.80±4.52	55.81±2.67
	2	102K	63.23±1.88	60.86±4.41	57.52±2.55
	4	156K	64.65±1.79	63.11±4.35	59.35±2.43
	8	265K	65.59±1.75	63.09±4.22	59.27±2.30

Table 10: Capacity-matched comparison on PPMI (%): all models use hidden dimension 64, batch size 16 and at most 500 epochs, with 10-fold cross-validation.

Model	Acc	Pre	F1
FFSNN	65.73±3.86	39.48±9.93	38.69±7.30
SRNN	64.73±2.53	44.46±6.08	39.99±6.29
LSNN	67.26±5.90	40.61±10.71	39.57±9.38
ASRNN	64.72±2.90	42.86±3.03	37.70±5.27
DH-SNN	61.85±2.86	40.42±12.68	36.07±9.25
Rhythm-SNN	64.21±2.91	42.22±8.66	36.03±6.01
S^2-Net	**67.62±5.21**	**49.30±8.83**	**41.51±7.19**

Specifically, extreme values of the coupling strength (K) reduce performance. Weak coupling ($K = 0.1$) fails to establish sufficient phase coherence, leading to the lowest F1-score (38.12%), while overly strong coupling ($K = 10$) forces premature global synchronization. With $K = 1.0$, precision reaches 49.30%, higher than at both boundary values.

Removing the time delay ($\tau = 0$) also lowers performance, which suggests that instantaneous coupling is less suited to resolving spatiotemporal patterns. Excessive delays (e.g., $\tau = 5$) also reduce the F1-score (to 39.06%), possibly because of "over-frustration" in the oscillator network. These results suggest that a moderate delay ($\tau = 2$) breaks phase symmetry without disrupting the temporal alignment of the gating.

Regarding the *oscillation dimension (D)*, S^2-Net is stable across dimensions (Table 11). The performance differences between $D = 1, 4$, and 8 are minimal, with $D = 1$ showing a slight edge in raw accuracy. Since all differences are within one standard deviation, we keep $D = 4$ as the default for consistency across datasets.

Table 11: Sensitivity analysis of S^2-Net hyperparameters on the PPMI dataset.

Variable	K	τ	D	**Acc (%)**	**Pre (%)**	**F1 (%)**
K	0.1	2	4	66.77 ± 3.66	46.03 ± 8.22	38.12 ± 6.78
	10	2	4	65.92 ± 3.76	46.54 ± 3.64	40.24 ± 4.94
τ	1.0	0	4	67.28 ± 4.04	46.25 ± 3.89	40.34 ± 6.02
	1.0	5	4	66.93 ± 3.55	46.23 ± 8.32	39.06 ± 6.37
D	1.0	2	1	67.76 ± 4.72	47.17 ± 4.61	41.86 ± 6.23
	1.0	2	8	67.43 ± 4.45	47.26 ± 3.68	39.44 ± 4.95
Default	**1.0**	**2**	**4**	**67.62 ± 5.21**	**49.30 ± 8.83**	**41.51 ± 7.19**

A.6 Implementation Details

We use the NVIDIA H100 NVL (8GPUs,94G/per) to conduct the involved experiments. We set time delay $\tau = 2$, oscillation dimension $D = 4$. The detailed hyperparameter settings are shown in Table 12. Table 13 shows the number of parameters and inference time of each model. S^2-Net has a model size of 0.22 MB, comparable to Rhythm-SNN and much smaller than recurrent baselines such as

SRNN and LSNN. Regarding inference speed, S^2-Net has the lowest throughput among the compared models (2.79 samples/s), because the oscillator dynamics are integrated sequentially at every time step.

Table 12: Hyperparameter settings for different models.

Model	FFSNN	SRNN	LSNN	ASRNN	DH-SNN	Rhythm-SNN	**S^2-Net**
Optimizer	Adam	Adam	Adam	Adam	Adam	Adam	Adam
Learning rate	10^{-2}	10^{-2}	10^{-2}	10^{-2}	10^{-2}	10^{-2}	10^{-3}
Batch size	256	64	256	64	16	16	16
Epochs	250	250	250	250	500	500	500
LR Schedule	StepLR (100, 0.75)	StepLR (100, 0.75)	StepLR (100, 0.75)	StepLR (100, 0.75)	StepLR (100, 0.75)	StepLR (100, 0.75)	StepLR (100, 0.75)
Hidden dim	1024	1024	1024	1024	1024	1024	64

Table 13: Number of network parameters and inference time on HCP-WM dataset ($N = 116$ and $T = 405$).

Model	FFSNN	SRNN	LSNN	ASRNN	DH-SNN	Rhythm-SNN	**S^2-Net**
# parameters (MB)	0.52	4.52	4.52	4.53	0.18	0.12	**0.22**
infer time (samples/s)	9.68	6.05	3.76	3.94	5.50	8.51	**2.79**

Accessibility. All datasets used in this study are publicly available and can be accessed from their respective sources. HCP-A is available at `https://www.humanconnectome.org/study/hcp-lifespan-aging/data-releases`. HCP-YA and HCP-WM are available at `https://www.humanconnectome.org/study/hcp-young-adult/data-releases`. UKB dataset is available at `https://www.ukbiobank.ac.uk/`. ADNI, PPMI and NIFD can be found in `https://ida.loni.usc.edu/login.jsp`. The S-MNIST and PS-MNIST datasets are derived from the MNIST dataset, which is available at `http://yann.lecun.com/exdb/mnist/`. The SHD dataset can be accessed at `https://zenkelab.org/resources/spiking-heidelberg-datasets-shd/`. The ECG dataset is publicly available from PhysioNet at `https://physionet.org/content/qtdb/1.0.0/`. The Google Speech Commands (GSC) dataset can be obtained from `https://www.tensorflow.org/datasets/catalog/speech_commands`.

Preprocessing and experimental protocols follow Yan et al. (2025). Access to HCP, UK Biobank, ADNI, PPMI and NIFD is governed by their data use agreements, which we followed; MNIST, SHD, ECG (PhysioNet QTDB) and GSC are used under their published terms.

A.7 Impact Statement

This work introduces a biologically inspired architecture (S^2-Net) motivated by cortical oscillatory synchronization. Spiking models may contribute to energy-efficient computation on neuromorphic hardware, although we have not measured energy consumption in this work. Our experiments on neurodegenerative disease datasets (ADNI, PPMI and NIFD) suggest potential for analyzing spatiotemporal biomedical data; any diagnostic use would require clinical validation, which this work does not provide.

The neuroimaging data used here are de-identified and were accessed under the data use agreements of the respective studies (Appendix A.6).

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: Please check Abstract and Introduction - Sec. 1.

   Guidelines:

   - The answer [N/A] means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A [No] or [N/A] answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: Please check the Limitations paragraph in Sec. 4.

   Guidelines:

   - The answer [N/A] means that the paper has no limitation while the answer [No] means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

   Answer: [Yes]

Justification: Please check Sec. A.2.

Guidelines:

- The answer [N/A] means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

   Answer: [Yes]

   Justification: Please check Sec. 3 - Experiments.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - If the paper includes experiments, a [No] answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
   - If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
   - Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
   - While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
     (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
     (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
     (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
     (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

Answer: [Yes]

Justification: Code, configurations and instructions are available at GitHub; the datasets are publicly available under their data use agreements (Appendix A.6).

Guidelines:

- The answer [N/A] means that paper does not include experiments requiring code.
- Please see the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
- While we encourage the release of code and data, we understand that this might not be possible, so [No] is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
- The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
- The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
- The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
- At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
- Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer) necessary to understand the results?

   Answer: [Yes]

   Justification: Please check Sec. A.6 - Implementation Details.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
   - The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [Yes]

   Justification: Tables 1, 4 and 5 report mean $\pm$ standard deviation over 10 cross-validation folds, and significance in Table 1 is assessed with paired t-tests across folds. Table 2 reports accuracy without error bars, following prior work (Yan et al., 2025).

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The authors should answer [Yes] if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
   - The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
   - The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
   - The assumptions made should be given (e.g., Normally distributed errors).

- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g., negative error rates).
- If error bars are reported in tables or plots, the authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [Yes]

   Justification: Hardware (NVIDIA H100 NVL GPUs) is described in Appendix A.6, and model size and inference throughput are reported in Table 13.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: [Yes]

   Justification: Please check the whole paper.

   Guidelines:

   - The answer [N/A] means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer [No], they should explain the special circumstances that require a deviation from the Code of Ethics.
   - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [Yes]

    Justification: Please check Sec. A.7 - Impact Statement.

    Guidelines:

    - The answer [N/A] means that there is no societal impact of the work performed.
    - If the authors answer [N/A] or [No], they should explain why their work has no societal impact or why the paper does not address societal impact.
    - Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.

- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate Deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

    Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pre-trained language models, image generators, or scraped datasets)?

    Answer: [N/A]

    Justification: The paper poses no such risks

    Guidelines:

    - The answer [N/A] means that the paper poses no such risks.
    - Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
    - Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
    - We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

    Answer: [Yes]

    Justification: All datasets are cited. HCP, UK Biobank, ADNI, PPMI and NIFD are used under their respective data use agreements; MNIST, SHD, ECG (PhysioNet QTDB) and GSC are used under their published terms. Access links are given in Appendix A.6.

    Guidelines:

    - The answer [N/A] means that the paper does not use existing assets.
    - The authors should cite the original paper that produced the code package or dataset.
    - The authors should state which version of the asset is used and, if possible, include a URL.
    - The name of the license (e.g., CC-BY 4.0) should be included for each asset.
    - For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
    - If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.

- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [Yes]

    Justification: We release our code under an open-source license with documentation at `https://github.com/Dandy5721/S2Net`.

    Guidelines:

    - The answer [N/A] means that the paper does not release new assets.
    - Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
    - The paper should discuss whether and how consent was obtained from people whose asset is used.
    - At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: [N/A]

    Justification: The paper does not involve crowdsourcing nor research with human subjects.

    Guidelines:

    - The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
    - Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
    - According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: [N/A]

    Justification: The paper does not involve crowdsourcing nor research with human subjects.

    Guidelines:

    - The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
    - Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
    - We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.

- For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

    Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does *not* impact the core methodology, scientific rigor, or originality of the research, declaration is not required.

    Answer: [N/A]

    Justification: The core method development in this research does not involve LLMs as any important, original, or non-standard components.

    Guidelines:

    - The answer [N/A] means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
    - Please refer to our LLM policy in the NeurIPS handbook for what should or should not be described.