Symmetry-Aligned Priors Can Dominate Spike Activity in Neural Circuit Inference

Seungwon Yu [1] **Jaeho Yang** [1] **Kijung Yoon** [1 2]

Abstract

Inferring synaptic weights from neural activity is commonly framed as a data-driven problem, yet inferred connectivity can become largely independent of spike inputs even when models are trained on them. Using a graph-based inference framework, we show that positional encoding alone can recover stable connectivity structure, with inferred weights remaining consistent when spike inputs are removed at test time. We interpret this as evidence that symmetry-aligned geometric priors can dominate activity-dependent evidence in connectivity inference. These findings highlight the strong influence of inductive bias and suggest that neural circuit inference should be understood as an interaction between neural data and geometric priors.

1 Introduction

Inferring synaptic weights from neural population activity is a central problem in computational neuroscience. This task is typically framed as a data-driven problem in which statistical dependencies in spike trains reveal underlying circuit structure (Pillow et al., 2008; Schneidman et al., 2006; Paninski, 2004). Recent approaches based on message-passing neural networks have advanced this direction by jointly learning latent weights and predicting neural activity in a self-supervised manner (Yoon, 2025). These models assume that connectivity can be recovered from temporal patterns in spike data, even in strongly recurrent networks. In this work, "weights" refer to inferred synaptic weights W, treated as latent variables produced by the model rather than its trainable parameters.

We revisit this assumption from the perspective of the latent space of inferred synaptic weights. Specifically, we study whether inferred connectivity genuinely reflects information extracted from neural activity, or whether it can instead arise from structural priors encoded in the model. This question becomes particularly important when extending static connectivity inference frameworks to dynamic settings, where connectivity is estimated over time.

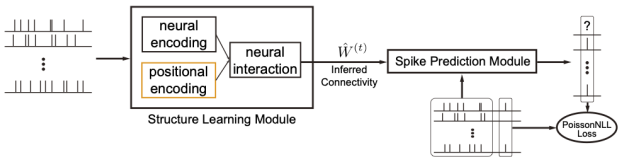

Figure 1. Overview of the graph-based connectivity inference model (Yoon, 2025). Neural encoding represents spike-driven features, while positional encoding (orange box) provides a symmetry-aligned geometric prior introduced here, organizing the space of inferred connectivity.

We begin by examining a necessary condition for dynamic inference: whether a model can consistently recover a time-invariant connectivity structure even when applied at every time step. Using the same ring attractor benchmark with a Mexican-hat weight profile as in prior work (Das & Fiete, 2020; Yoon, 2025), we find that a naïve extension of the model to dynamic inference fails to recover the true structure. Surprisingly, however, introducing positional encoding (PE) enables consistent recovery of the underlying connectivity across all time steps.

Further analysis shows that this recovery is not driven by spike data. The model instead exploits PE as a geometric prior, reconstructing connectivity even when spike inputs are removed. We interpret this phenomenon as evidence that symmetry-aligned geometric priors can dominate activity-dependent evidence in connectivity inference. This finding highlights a fundamental tension between inductive bias and data-driven inference (Richards et al., 2019; Dorrell et al., 2023; Battaglia et al., 2018) and suggests that neural circuit inference should be understood as learning within a symmetry-constrained weight space (Han et al., 2026; Bronstein et al., 2021).

[1]Department of Electronic Engineering, Hanyang University, Seoul, South Korea [2]Department of Artificial Intelligence, Hanyang University, Seoul, South Korea. Correspondence to: Kijung Yoon <kiyoon@hanyang.ac.kr>.

Workshop on Weight-Space Symmetries, held in conjunction with the 43rd International Conference on Machine Learning, Seoul, South Korea. 2026. Copyright 2026 by the author(s).

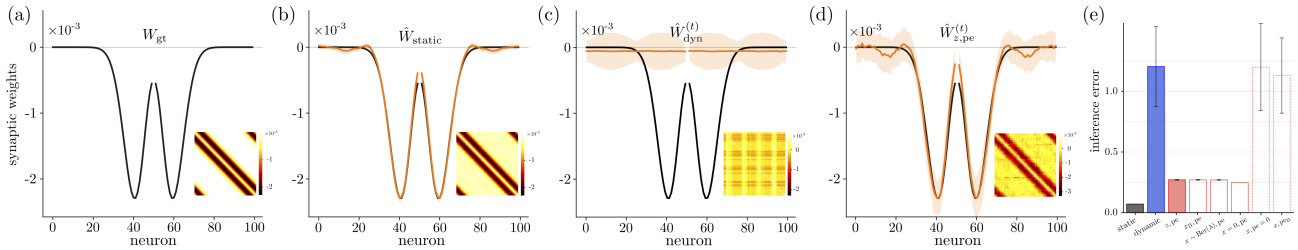

Figure 2. Connectivity inference across static and dynamic regimes. (a) Ground-truth Mexican-hat weight profile (black), corresponding to the connectivity profile of every neuron in the network, and weight matrix (inset). (b) Static inference (orange) recovers the structure. (c) Dynamic inference using a single time-step snapshot fails due to limited temporal context. (d) Dynamic inference with positional encoding restores consistent structure. (e) Weight inference error across settings. Here, (z, pe) denotes the baseline in which PE is concatenated with neural embeddings; x_Π and pe_Π denote permutation of neural activity and PE across neurons; $x \sim \mathrm{Ber}(\lambda)$ denotes replacement of spike inputs with Bernoulli random samples; and $x = 0$, $pe = 0$ denote replacement with zeros. Dynamic results are averaged over 10K time steps with standard error, while static inference and the zero-spike condition are shown as single estimates without error bars.

2 Problem Setup

We consider the problem of inferring neural connectivity from spike train data generated by a recurrent network with fixed structure. The underlying generative model is a ring attractor network with a Mexican-hat weight profile (Figure 2a), which produces structured, spatially organized neural activity (Zhang, 1996; Seung, 1996; 1998; Das & Fiete, 2020). The goal is to recover the pairwise synaptic weights $W \in \mathbb{R}^{N\times N}$ from observed spike trains $x \in \{0,1\}^{N\times T}$.

The inference model comprises two components: a structure learning module and a spike prediction module (Figure 1). The structure learning module encodes each neuron's spike activity into an embedding z_i and infers pairwise synaptic weights,

$$z_i = f_{\text{enc}}(x_i), \quad w_{ij} = f_{\text{edge}}([z_i, z_j]).$$

The resulting weight matrix $\hat{W}$ is then used by the spike prediction module, which combines it with neuronal spike histories to predict subsequent activity through message passing across neurons.

Although the original formulation assumes time-invariant connectivity, we extend it to a dynamic setting by estimating $W^{(t)}$ at each time step using a sliding window over spike history. This formulation allows us to evaluate whether the model can consistently recover a fixed connectivity structure across time, which we view as a necessary condition for reliable inference of time-varying connectivity.

3 Experiments

3.1 Static Inference

As a reference point, we first reproduce static connectivity inference using the original model formulation. In the static setting, each neuron's spike train $x_i \in \{0,1\}^T$ spans the full recording, and the structure learning module maps it to an embedding

$$z_i = f_{\text{enc}}(x_i),$$

where $f_{\text{enc}}(\cdot)$ is a temporal feature extractor composed of a 1D convolution followed by a fully connected projection (Yoon, 2025; Park et al., 2022). The convolution aggregates spike activity over a temporal window matched to the synaptic timescale, and the resulting features are pooled across time to produce a compact representation. Pairwise weights are then inferred as $w_{ij} = f_{\text{edge}}([z_i, z_j])$. With long recordings, this formulation reliably recovers the Mexican-hat connectivity (Figure 2b).

3.2 Dynamic Inference with Sliding Windows

We then extend the same architecture to a dynamic setting by applying it to successive sliding windows of the neural recording, treating each window as a partial observation of the full dataset. At each time step,

$$z_i^{(t)} = f_{\text{enc}}(x_i^{(t-\ell+1:t)}),$$

and connectivity is inferred as $\hat{W}^{(t)}$. Although the functional form of f_{enc} is unchanged, the input now consists of short spike segments of length ℓ. The key difference is the effective sample size. In the static setting, each embedding is computed from long spike trains with millions of time steps, enabling stable estimation of cross-neuronal correlations. In the dynamic setting, embeddings are computed from short windows, which sharply reduces the number of spikes contributing to each estimate and limits the reliability of the correlation structure available to the model.

Consistent with this limitation, the model fails to recover the Mexican-hat profile in the dynamic regime (Figure 2c). This suggests that the structure learning module, while expressive, is data-limited: when temporal context is insufficient, it cannot reliably infer weights from spike activity alone. The success of static inference therefore depends on large temporal sample sizes. Once this condition is violated, the same architecture fails, motivating the need for additional inductive bias to stabilize connectivity estimation.

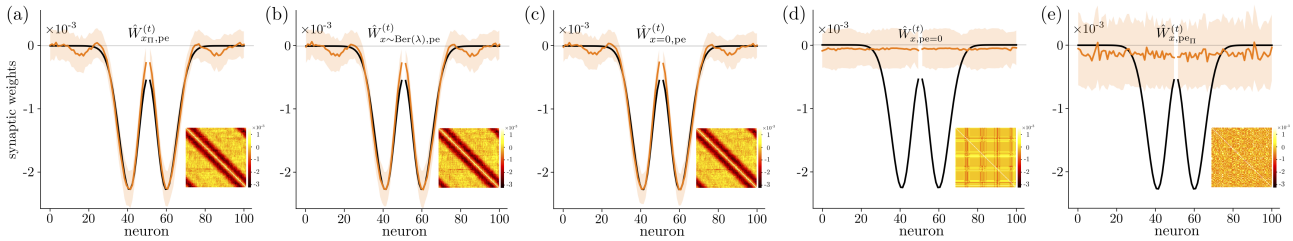

Figure 3. Controlled perturbation experiments isolating the roles of spike inputs and positional encoding. (a–e) Inferred connectivity from the dynamic setting, shown as single time-step snapshots. (a–c) Spike perturbations with PE fixed: (a) permuting neuron indices, (b) replacing spikes with independent Bernoulli samples, and (c) zeroing spike inputs all preserve the inferred connectivity. (d–e) PE perturbations with spike inputs fixed: (d) removing PE and (e) permuting PE disrupt connectivity, with permutation yielding the largest degradation.

3.3 Effect of Positional Encoding

To improve dynamic inference, we augment the structure learning module with a learnable PE (Figure 1) (Dwivedi et al., 2022; Wang et al., 2022). Each neuron i is assigned a vector $p_i \in \mathbb{R}^N$, with encoding dimensionality set equal to the number of neurons. The encoding vectors are initialized independently from a standard normal distribution, $p_i \sim \mathcal{N}(0, I)$, and optimized jointly with the rest of the model during training. The embedding is formed as

$$z_i' = [z_i, p_i]$$

and used to infer connectivity as before[1].

With this modification, the model consistently recovers the Mexican-hat weight profile across time steps (Figure 2d). In contrast to the failure without PE, the inferred weights are stable and exhibit the expected spatial structure, although the overall inference quality remains below that of static inference (Figure 2e). This suggests that positional information provides a strong inductive bias that compensates for limited temporal context.

This result raises a fundamental question: does PE simply regularize the inference problem, or does it change what the model relies on? The ability to recover connectivity despite limited spike data suggests that the model may exploit positional structure rather than extract information from neural activity. We therefore perform perturbation studies to identify the information source used by the learned mapping.

3.4 Perturbation Analysis

To isolate the respective contributions of spike inputs and positional encoding, we perform controlled perturbation experiments that selectively modify each source of information. All perturbations are conducted *without retraining the model*. We use the trained model as-is and alter only the inputs to the structure learning module at inference time, thereby probing what information the learned mapping actually relies on.

Spike Perturbations with PE Fixed We first disrupt spike-based information while keeping positional encoding fixed. In the first perturbation, neuron indices in the spike data are randomly permuted, so that each neuron receives another neuron's spike train while retaining its original positional encoding p_i. This operation preserves overall spike statistics but breaks the correspondence between spike activity and neuron identity. Despite this mismatch, the inferred connectivity remains unchanged relative to the baseline with PE (Figures 3a & 2e), indicating that the model is invariant to spike identity when PE is present.

We then remove statistical structure more directly by replacing the observed spikes with independent random samples drawn from a Bernoulli distribution, $x_i^{(t)} \sim \mathrm{Ber}(\lambda)$, where λ is set to the average firing rate of the original spike trains. Although this perturbation destroys all temporal and cross-neuronal correlations, the model recovers a connectivity structure similar to the baseline with PE (Figures 3b & 2e).

Finally, we replace the spike inputs with zeros while keeping PE fixed. The inferred weights remain consistent with the preceding perturbations (Figure 3c), indicating reliance on PE alone. The weight inference error is slightly lower in this setting (Figure 2e). Together, these observations imply that spike-derived information is not necessary for accurate inference in the presence of PE.

PE Perturbations with Spike Inputs Fixed We next fix the spike inputs and perturb the positional encoding. Removing positional information entirely by setting $p_i = 0$ for all neurons prevents the model from recovering meaningful connectivity (Figure 3d), with inference error comparable to dynamic inference without PE (Figures 2c & 2e). This indicates that, in the absence of PE, the model reverts to the data-limited regime in which spike activity alone is insufficient for reliable connectivity estimation.

Permuting positional encodings across neurons assigns each neuron a mismatched encoding vector while keeping spike

[1] Alternative formulations of PE are provided in the Appendix A and do not affect the main results or interpretations.

inputs unchanged. This perturbation degrades the inferred connectivity (Figure 3e), with inference error comparable to both the zero-PE condition and dynamic inference without PE (Figure 2e), within the measured standard error of the mean. Incorrect geometric information is therefore at least as detrimental as the absence of positional information. Overall, these findings show that connectivity inference depends critically on PE, whereas spike activity plays a negligible role when PE is available.

4 Geometric Priors in Weight-Space Learning

We interpret these results through the lens of symmetry and geometric structure in synaptic weight space.

4.1 Symmetry Structure of the Ring Attractor

The Mexican-hat connectivity in ring attractor networks is invariant under rotation: synaptic weights depend only on the relative angular distance between neurons,

$$w_{ij} = f(d(i, j)).$$

Consequently, the connectivity matrix possesses a low-dimensional geometric structure. This is evident from the singular value decomposition (SVD) of the weight matrix, where the cumulative explained energy saturates rapidly and a small number of singular modes capture most of the variance in the true connectivity (Figure 4, Appendix B). This structure is independent of absolute neuronal identity and instead reflects the underlying rotational symmetry of the network. If the learned positional encodings p_i organize neurons within a latent coordinate system that reflects this geometry, then the edge function f_{edge} can infer connectivity primarily from the pair (p_i, p_j). In this case, the mapping effectively reduces to

$$(p_i, p_j) \mapsto w_{ij}.$$

From this perspective, PEs provide a coordinate system in which the geometric structure of the network can be represented. Rather than estimating connectivity directly from noisy spike correlations within each short temporal window, the model can exploit regularities encoded in the positional representations. When these representations capture the latent organization of neurons along the ring manifold, connectivity can be reconstructed from relative relationships between coordinates.

Our perturbation experiments are consistent with this interpretation. Connectivity estimates remain largely unchanged when spike activity is randomized or removed, but deteriorate when PEs are perturbed. These results suggest that, in the dynamic regime, connectivity estimates are determined primarily by the coordinate structure represented by the PEs

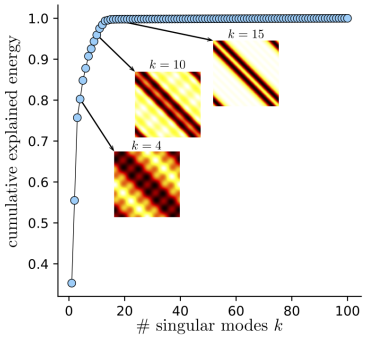

Figure 4. Cumulative explained energy from singular values of the true weight matrix W, showing rapid saturation, with a small number of singular modes capturing most of the variance; insets illustrate reconstructions using leading singular modes ($k = 4, 10, 15$).

rather than by information extracted from short spike windows. We therefore hypothesize that PEs act as a geometric prior, providing a latent coordinate system that enables the model to reconstruct connectivity from structural regularities when spike-derived information is insufficient.

These results do not imply that spike activity contains no information about connectivity. Instead, they indicate that in the present benchmark, the symmetry structure of the connectivity matrix provides a sufficiently strong constraint for the model to recover connectivity without relying substantially on activity-derived information. Whether the same phenomenon persists in less symmetric networks remains an open question.

4.2 PE-Only Structure Learning

To test whether spike-derived features are required once positional information is available, we *retrained the model using positional encodings alone*, removing spike activity from the structure learning module. In this setting, neuron embeddings were given by $z_i = p_i$, and connectivity was inferred directly through

$$w_{ij} = f_{\text{edge}}([p_i, p_j]).$$

Although spike-derived features were removed from the structure learning input, the model was still trained through the downstream spike prediction objective, so neural activity continued to provide the training signal for optimizing the coordinate-to-connectivity mapping.

The model continued to recover the Mexican-hat connectivity profile (Figures 7a-7b). Combined with the perturbation experiments, this result indicates that PEs alone contain sufficient information for connectivity reconstruction and that the model does not primarily rely on correlations extracted from short spike windows. Instead, the edge network can learn a mapping from pairs of neuron-specific coordinates to synaptic weights, effectively reconstructing the distance-dependent connectivity rule from the geometry encoded in PE.

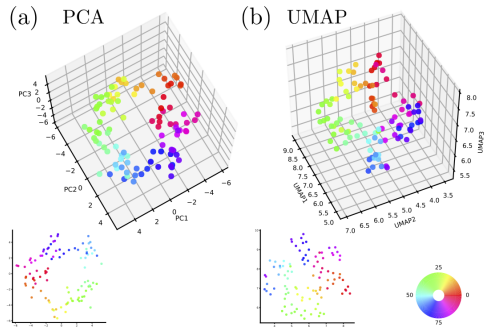

Figure 5. Visualization of positional encoding vectors learned in the PE-only training setup. PE vectors from all 100 neurons are projected into three-dimensional (top) and two-dimensional (bottom) spaces using (a) PCA and (b) UMAP. Each dot represents a PE vector and is color-coded according to neuron index (inset).

We further examined the learned positional encodings by projecting them with PCA and UMAP (McInnes et al., 2018). Across both visualization methods, the embeddings formed a ring-like organization (Figure 5), supporting the interpretation that learned PE captures a circular latent coordinate system aligned with the ring attractor geometry.

This observation connects to manifold-based analyses of head-direction circuits, where unsupervised characterization of population activity revealed a one-dimensional ring manifold corresponding to the encoded angular variable (Chaudhuri et al., 2019). Whereas those studies identify ring topology directly from neural population states, our analysis reveals a related ring-like organization in the learned PE space. Thus, circular geometry can emerge not only in population activity, but also in model-internal representations trained to predict future population activity and reconstruct connectivity. These findings support the coordinate-system hypothesis: PE may serve as a learned scaffold for representing network structure, bridging activity-manifold structure and model-internal geometric priors.

4.3 Fixed PE as a Coordinate System

If the coordinate-system hypothesis is correct, then connectivity recovery should not depend on the specific learned values of the PEs. Instead, any fixed coordinate system that consistently distinguishes and orders neurons should provide a sufficient reference frame for the edge network. To test this prediction, we trained the model from scratch using fixed sinusoidal positional encodings (Vaswani et al., 2017) in place of trainable neuron embeddings and repeated the dynamic inference procedure.

The model successfully recovered the Mexican-hat connectivity pattern even though the PEs were not learned (Figure 7c). This result suggests that the benefit of PE does not arise from learning a dataset-specific set of coordinates. Rather, fixed sinusoidal encodings provide a stable reference frame in which neurons can be consistently compared, allowing the edge network to represent structured relationships between neuron pairs.

Together, the learnable PE-only and fixed-sinusoidal PE experiments suggest that the critical factor is not the specific learned embedding values, but the availability of a stable neuron-specific coordinate system. Such a coordinate system allows the edge network to represent the distance-dependent Mexican-hat rule even when spike-derived evidence is sparse or absent.

5 Conclusion

We investigated dynamic connectivity inference in ring attractor networks using a recently proposed structure-learning framework. The original model successfully recovered connectivity from long recordings, but performance deteriorated when inference was performed on short sliding windows, indicating that the available spike-derived information was insufficient for reliable connectivity estimation.

Introducing positional encodings substantially improved connectivity recovery in the dynamic setting. Through a series of perturbation analyses, we found that inferred connectivity was far more sensitive to disruptions of positional encodings than to perturbations of spike activity. Successful inference remained possible when spike inputs were removed entirely from the structure learning module and when learned positional encodings were replaced by fixed sinusoidal representations. These results suggest that positional encodings primarily provide a coordinate system that enables the model to represent and exploit geometric regularities in the connectivity structure.

These findings do not imply that neural activity is unnecessary. Instead, they reveal a distinction between structure learning and activity prediction. In our experiments, positional information was sufficient for recovering connectivity within the structure learning module, whereas the spike prediction module continued to rely on neural activity to model network dynamics. Successful connectivity recovery therefore does not necessarily indicate that a model has extracted connectivity from data (Geirhos et al., 2020); it may instead reflect alignment between the model's inductive biases and the geometry of the underlying connectivity structure.

More broadly, our findings underscore the role of geometric priors in weight-space learning (Han et al., 2026). For structured systems such as ring attractor networks, connectivity inference may depend as much on the representation of the solution space as on the quantity of available neural data. Extending these ideas to settings with genuinely time-varying connectivity remains an important direction for future work and will likely require more principled integration of spike-derived information and geometric structure.

Acknowledgements

We thank the anonymous reviewers for their thoughtful feedback and constructive suggestions, which helped improve the clarity and presentation of the paper. K.Y. is supported in part by the National Research Foundation of Korea (NRF) grant (No. RS-2024-00337092), and the Institute of Information & communications Technology Planning & Evaluation (IITP) grants (No. RS-2020-II201373, Artificial Intelligence Graduate School Program; No. IITP-(2025)-RS-2023-00253914, Artificial Intelligence Semiconductor Support Program (Hanyang University)) funded by the Korean government (MSIT).

References

Battaglia, P. W., Hamrick, J. B., Bapst, V., Sanchez-Gonzalez, A., Zambaldi, V., Malinowski, M., Tacchetti, A., Raposo, D., Santoro, A., Faulkner, R., et al. Relational inductive biases, deep learning, and graph networks. *arXiv preprint arXiv:1806.01261*, 2018.

Bronstein, M. M., Bruna, J., Cohen, T., and Veličković, P. Geometric deep learning: Grids, groups, graphs, geodesics, and gauges. *arXiv preprint arXiv:2104.13478*, 2021.

Chaudhuri, R., Gerçek, B., Pandey, B., Peyrache, A., and Fiete, I. The intrinsic attractor manifold and population dynamics of a canonical cognitive circuit across waking and sleep. *Nature Neuroscience*, 22(9):1512–1520, 2019.

Das, A. and Fiete, I. R. Systematic errors in connectivity inferred from activity in strongly recurrent networks. *Nature Neuroscience*, 23(10):1286–1296, 2020.

Dorrell, W., Yuffa, M., and Latham, P. E. Meta-learning the inductive bias of simple neural circuits. In *International Conference on Machine Learning*, pp. 8389–8402, 2023.

Dwivedi, V. P., Luu, A. T., Laurent, T., Bengio, Y., and Bresson, X. Graph neural networks with learnable structural and positional representations. In *International Conference on Learning Representations*, 2022.

Geirhos, R., Jacobsen, J.-H., Michaelis, C., Zemel, R., Brendel, W., Bethge, M., and Wichmann, F. A. Shortcut learning in deep neural networks. *Nature Machine Intelligence*, 2(11):665–673, 2020.

Han, X., Wang, Z., Zhao, B., Zhang, B., Li, J., Borth, D., Yu, R., Maron, H., Ye, Y., Yin, L., et al. A survey of weight space learning: Understanding, representation, and generation. *arXiv preprint arXiv:2603.10090*, 2026.

McInnes, L., Healy, J., and Melville, J. Umap: Uniform manifold approximation and projection for dimension reduction. *arXiv preprint arXiv:1802.03426*, 2018.

Paninski, L. Maximum likelihood estimation of cascade point-process neural encoding models. *Network: Computation in Neural Systems*, 15(4):243, 2004.

Park, T., Kim, J., Kang, D., and Yoon, K. Graph neural networks for connectivity inference in spatially patterned neural responses. In *NeurIPS 2022 Workshop on Symmetry and Geometry in Neural Representations*, 2022.

Pillow, J. W., Shlens, J., Paninski, L., Sher, A., Litke, A. M., Chichilnisky, E., and Simoncelli, E. P. Spatio-temporal correlations and visual signalling in a complete neuronal population. *Nature*, 454(7207):995–999, 2008.

Richards, B. A., Lillicrap, T. P., Beaudoin, P., Bengio, Y., Bogacz, R., Christensen, A., Clopath, C., Costa, R. P., de Berker, A., Ganguli, S., et al. A deep learning framework for neuroscience. *Nature Neuroscience*, 22(11): 1761–1770, 2019.

Schneidman, E., Berry, M. J., Segev, R., and Bialek, W. Weak pairwise correlations imply strongly correlated network states in a neural population. *Nature*, 440(7087): 1007–1012, 2006.

Seung, H. S. How the brain keeps the eyes still. *Proceedings of the National Academy of Sciences*, 93(23): 13339–13344, 1996.

Seung, H. S. Continuous attractors and oculomotor control. *Neural Networks*, 11(7-8):1253–1258, 1998.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., and Polosukhin, I. Attention is all you need. *Advances in Neural Information Processing Systems*, 30, 2017.

Wang, H., Yin, H., Zhang, M., and Li, P. Equivariant and stable positional encoding for more powerful graph neural networks. In *International Conference on Learning Representations*, 2022.

Yoon, K. Self-supervised discovery of neural circuits in spatially patterned neural responses with graph neural networks. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*, 2025.

Zhang, K. Representation of spatial orientation by the intrinsic dynamics of the head-direction cell ensemble: a theory. *Journal of Neuroscience*, 16(6):2112–2126, 1996.

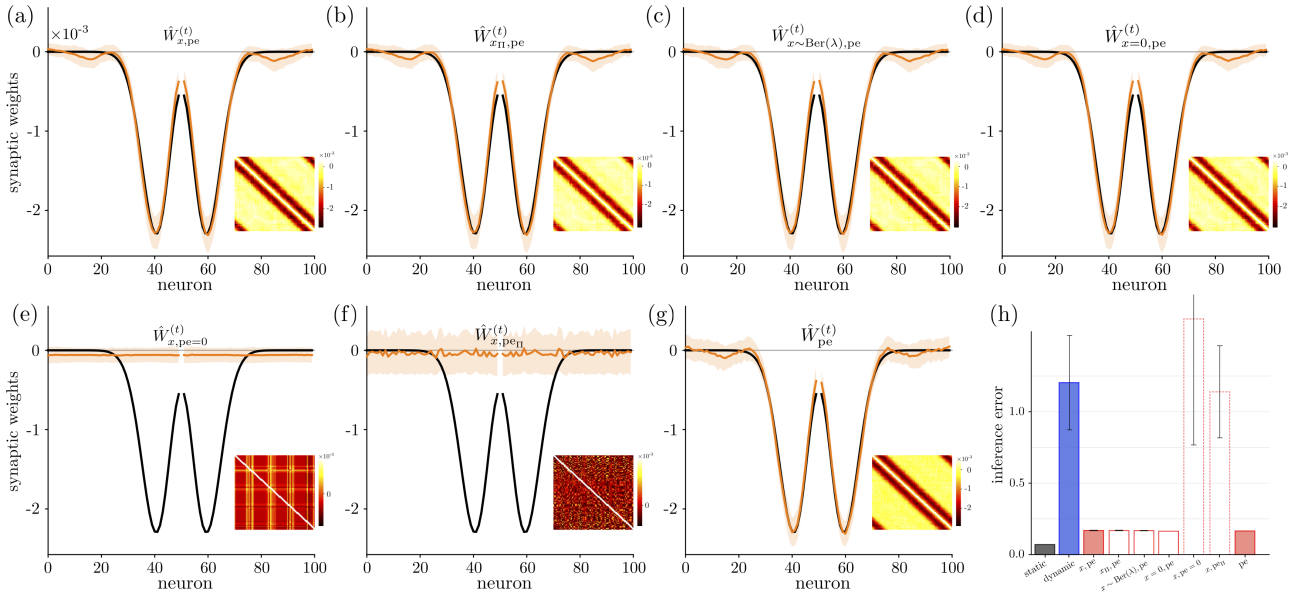

Figure 6. (a–f) Inferred weight profiles and corresponding weight matrices across settings when positional encoding is concatenated directly with spike inputs, denoted as (x, pe). (g) Inferred connectivity from the *pe* condition, in which the model is trained using positional encoding alone without spike inputs. (h) Weight inference error across settings. All perturbation notation is consistent with those used in Figure 2e.

A Alternative Formulation of Positional Encoding

In the main text, positional encoding is incorporated by concatenating p_i with the neural encoding z_i. As an alternative formulation, we instead concatenate positional encoding directly with the spike inputs, training the model using $[x_i, p_i]$ as input to the structure learning module. In this setting, positional information is introduced prior to feature extraction, rather than at the embedding level.

We evaluate this variant using the same experimental protocol and perturbation analysis described in the main text. The resulting inference errors exhibit a distribution similar to that shown in Figure 2e, indicating that the qualitative behavior of the model remains unchanged under this modification (Figure 6). In particular, the model configuration denoted as *pe*, in which the model is trained using positional encoding alone without spike inputs, achieves inference error comparable to the counterpart trained with spike activity. This result shows that the dominance of positional encoding is not specific to the embedding-level concatenation used in the main text, but persists under an alternative implementation. The PE-only training condition introduced in Figure 6 serves as an additional control for evaluating the role of positional structure in connectivity inference. A more detailed analysis of this setting, including sensitivity to PE dimensionality and sample size, is provided in the following section.

Overall, these results confirm that alternative formulations of positional encoding lead to the same conclusions regarding the relative contributions of spike inputs and geometric structure.

B Singular Value Decomposition of Weight Matrix

To analyze the structure of the ground-truth connectivity matrix $W \in \mathbb{R}^{N\times N}$, we compute its singular value decomposition $W = U\Sigma V^\top$, where $\Sigma = \mathrm{diag}(\sigma_1, \ldots, \sigma_N)$ with $\sigma_1 \geq \cdots \geq \sigma_N \geq 0$. We quantify the effective dimensionality of W using the cumulative explained energy $E(k) = \frac{\sum_{i=1}^{k}\sigma_i^2}{\sum_{i=1}^{N}\sigma_i^2}$, which measures the fraction of total variance captured by the top-k singular modes. As shown in Figure 4, the cumulative energy rises rapidly, with a small number of modes capturing most of the variance. Consistently, low-rank reconstructions $W_k = \sum_{i=1}^{k}\sigma_i u_i v_i^\top$ visually match the ground-truth connectivity even for small k, indicating that W is well approximated by a low-dimensional subspace. This suggests that the connectivity matrix is well described by a low-dimensional structure rather than by N^2 independent degrees of freedom, supporting the interpretation that it lies in a constrained region of weight space shaped by underlying symmetry and structured interactions.

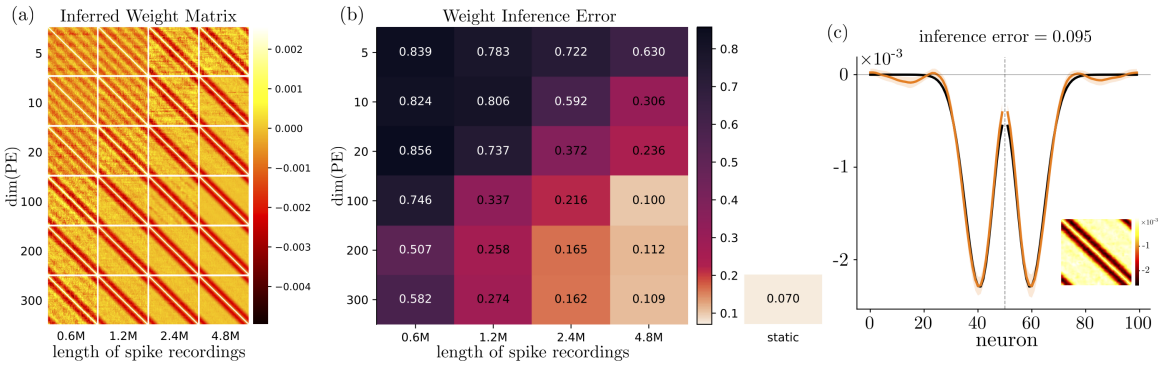

Figure 7. (a) Inferred weight matrices across PE-only training configurations. (b) Weight inference error of the PE-only training condition as a function of PE dimensionality and the total length of spike recordings used during training. Each entry reports the inference error obtained when the structure learning module is trained using positional encoding alone without spike-derived input. Increasing both PE dimension and training sample size generally improves inference quality, with the best-performing configuration achieved at $\dim(p_i) = 100$ and 4.8M training samples. The static inference error obtained using spike-driven training is shown separately for reference. The configurations in (b) are arranged in the same order as the inferred weight matrices shown in (a).

C Sensitivity of PE-Only Training to Encoding Dimension and Sample Size

To further characterize the extent of symmetry-induced degeneracy, we systematically evaluate the PE-only training condition across a range of positional encoding dimensionalities and training sample sizes. In these experiments, the structure learning module is trained without spike-derived input, using only positional encoding to infer connectivity. We vary the dimension of the PE as $\dim(p_i) \in \{5, 10, 20, 100, 200, 300\}$ and the total length of spike recordings used during training as $\{0.6\text{M}, 1.2\text{M}, 2.4\text{M}, 4.8\text{M}\}$ time steps, yielding 24 configurations in total.

The inferred connectivity exhibits strong dependence on both PE dimensionality and training sample size. When both quantities are small, the inferred weight matrices contain multiple off-diagonal stripe artifacts corresponding to distorted side dips in the inferred weight profiles, and the Mexican-hat structure is only weakly expressed (Figure 7a). As PE dimensionality and training sample size increase, these artifacts progressively diminish and the inferred connectivity more closely approaches the ground-truth Mexican-hat structure (Figure 7a). Consistent with these qualitative observations, increasing both PE dimension and training sample size generally improves inference quality (Figure 7b), indicating that the learned solution becomes increasingly constrained by positional structure aligned with the symmetry of the ground-truth connectivity. The best-performing configuration is obtained with $\dim(p_i) = 100$ and 4.8M time steps of training spike recordings, achieving a weight inference error of 0.10, approaching the static inference error of 0.07 obtained using spike-driven training (Figure 7b).

Importantly, the PE-only training condition should not be interpreted as a dynamic inference model. Because connectivity is inferred solely from fixed positional coordinates, this setting does not provide a mechanism for flexibly estimating time-varying connectivity from neural activity. Instead, it effectively reduces the problem to recovery of a static symmetry-aligned connectivity structure. The purpose of this experiment is therefore not to propose an alternative inference framework, but to probe the extent to which the learning objective admits solutions determined predominantly by positional structure.

The ability of PE alone to nearly match the performance of static spike-driven inference indicates that the connectivity estimation objective admits solutions largely constrained by symmetry-aligned geometric coordinates. In this regime, inferred connectivity can be recovered with only limited dependence on spike-derived evidence, despite the original objective being formulated as a data-driven inference problem.