# From Synapses to Dynamics: Obtaining Function from Structure in a Connectome Constrained Model of the Head Direction Circuit

**Sunny Duan***
Brain and Cognitive Sciences
Massachusetts Institute of Technology
sunnyd@mit.edu

**Ling Liang Dong***
Brain and Cognitive Sciences
Massachusetts Institute of Technology
loading@mit.edu

**Ila Fiete**
Brain and Cognitive Sciences
Massachusetts Institute of Technology
fiete@mit.edu

## Abstract

How precisely does circuit wiring specify function? This fundamental question is particularly relevant for modern neuroscience, as large-scale electron microscopy now enables the reconstruction of neural circuits at single-synapse resolution across many organisms. To interpret circuit function from such datasets, we must understand the extent to which the measured structure constrains dynamics. We investigate this question in the *Drosophila* head direction (HD) circuit, which maintains an internal heading estimate through attractor dynamics that integrate self-motion velocity cues. This circuit serves as a sensitive assay for functional specification: continuous attractor networks are theoretically known to require finely tuned wiring symmetries, whereas connectomes omit key cellular parameters such as synaptic gains, neuronal thresholds, and time constants, and reveal that biological wiring can be heterogeneous. We introduce a method that combines self-supervised and unsupervised learning objectives to estimate unknown parameters at the level of cell types, rather than individual neurons and synapses. Starting from the raw connectivity matrix, our approach recovers a network that exhibits continuous attractor dynamics and accurately integrates a range of velocity inputs, despite minimal parameter tuning on a connectome that notably departs from the symmetric regularity of an idealized ring attractor. We characterize how deviations from the original connectome shape the space of viable solutions. We also perform in-silico ablation experiments to probe the distinct functional roles of specific cell types in the circuit, demonstrating how connectome-derived structure, when augmented with minimal, biologically grounded tuning, can replicate known physiology and elucidate circuit function.

## 1 Introduction

Recent advances in large-scale electron microscopy have enabled the reconstruction of neural circuits at synapse-level resolution, producing dense connectomic datasets across organisms including *C. elegans* (Cook et al., 2019), *Drosophila* (Scheffer et al., 2020b; Dorkenwald et al., 2024), and mice (MICrONS Consortium, 2025). These datasets raise several fundamental questions: How precisely

*Equal contribution

39th Conference on Neural Information Processing Systems (NeurIPS 2025).

 https://doi.org/10.52202/085713-4273

can circuit function be determined from its wiring? Can we do so in the absence of key membrane and synaptic parameters such as cellular thresholds, gains, and time constants? Additionally, connectomes provide detailed synaptic connectivity maps but may contain potential errors from the data pipeline, including misalignment of electron microscopy sections and false positives or negatives in synapse detection Scheffer et al. (2020a). Thus, individual connectomes represent partial and sometimes noisy snapshots of the underlying biological networks that may fail to capture the necessary factors that play a role in shaping circuit behavior.

We explore these questions in the *Drosophila* head direction (HD) circuit, a canonical example of a continuous ring attractor network in biology (Kim et al., 2017; Green et al., 2017; Turner-Evans et al., 2017) that maintains an internal estimate of animal heading by integrating angular self-motion velocity inputs. Whereas theoretical studies of ring attractors (Skaggs et al., 1994; Zhang, 1996; Redish et al., 1996; Xie et al., 2002) show that they require finely tuned connectivity to support bump stability and smooth translation for velocity integration, we found that connectomic reconstructions of this circuit exhibit asymmetries and heterogeneities. Moreover, the *Drosophila* HD circuit involves $\sim 10$ distinct cell types — many more than are theorized to be necessary in theoretical ring attractor models — raising open questions about the necessity and roles of specific circuit components.

Previous models of the HD circuit (Chang et al., 2023; Stentiford et al., 2024) have relied on hand-designed circuit architectures that impose connectivity motifs, in particular circular symmetry, known from earlier work to support ring attractor dynamics (Skaggs et al., 1994; Zhang, 1996; Redish et al., 1996; Xie et al., 2002). While such models produce the expected network dynamics by design, they start with prior knowledge about circuit structure and the identities of the cell types involved in the core circuit, thus limiting opportunities for data-driven discovery. Here, we introduce a framework that bridges this gap: starting from the original measured connectivity matrix, we optimize only a small set of biologically grounded parameters at the cell-type level, such as cell-type-to-cell-type synaptic gains and cell type-shared neuronal biases and time constants. The model is trained using a self-supervised linear consistency loss that simply enforces the internal bump movement to be proportional to the input velocity, in addition to several regularization terms that ensure the model learns a non-trivial solution that remains close to the original connectome. Notably, this learning objective requires no neural activity recordings or behavioral labels, enabling inference of functional dynamics from only structural measurements.

Our trained model recovers high-fidelity continuous attractor dynamics and robustly performs angular velocity integration across a wide range of inputs. It also reproduces experimental findings including realistic bump width and bidirectional bump motion dependent on asymmetric input from the left and right P-EN subpopulations (Seelig and Jayaraman, 2015; Turner-Evans et al., 2017). We find that tuning parameters at the cell-type level achieves performance comparable to models with more parameters, while global parameter tying fails to recover attractor behavior—indicating that cell-type–specific tuning offers a practical tradeoff between model flexibility and parsimony, performing comparably to fully parameterized models while preserving biological interpretability. We also explore the space of viable solutions under different levels of simulated noise and characterize the diversity of model solutions as a function of their initial starting points on the loss landscape. Finally, we conduct in silico ablation experiments to probe the roles of different cell types.

**Key Contributions**

- We introduce a self-supervised connectome-constrained framework that recovers functional dynamics of integrator circuits directly from noisy connectome measurements
- We show that parameterization at the level of cell types is sufficient and necessary to recover integrating network dynamics
- We characterize how initial conditions shape the diversity of learned solutions.
- We perform in silico ablations on the trained networks, yielding novel biological insights on the functional roles of specific neuron classes in the Drosophila HD circuit.

## 2 Background

### 2.1 Continuous Attractor Models

Continuous attractor networks encode analog variables, such as orientation or position, via localized bumps of persistent activity that shift smoothly across the neural population. These dynamics require precise tuning of recurrent connectivity, typically balancing local excitation and broad inhibition (Khona and Fiete, 2022). Even minor deviations in synaptic strength or timing can cause bump instability, drift, or collapse. This sensitivity makes such networks a strong testbed for evaluating the extent to which connectomes alone can specify circuit dynamics.

### 2.2 Drosophila HD Circuit

The Drosophila HD circuit resides in the central complex and comprises several cell types (populations of cells that exhibit shared morphological, genetic, and connectivity profiles) thought to implement a ring attractor (Kim et al., 2017; Turner-Evans et al., 2020). E-PG neurons in the ellipsoid body (EB) are connected in an anatomical ring-like structure and comprise the core set of cells that track animal heading via a localized bump in population activity (Seelig and Jayaraman, 2015). P-EN neurons convey angular velocity signals: left and right P-EN neurons asymmetrically project to E-PG neurons to shift the bump based on the directional velocity, and are further subdivided into P-ENa and P-ENb subpopulations whose differential roles in the circuit have yet to be ascertained (Turner-Evans et al., 2017). Inhibitory inputs from Delta7 and ring neurons are believed to normalize and gate activity (Green et al., 2017; Hulse et al., 2023), while P-EG neurons are thought to contribute to bump stabilization (Pisokas et al., 2020), although this has not yet been confirmed experimentally. These cell types outnumber those in idealized ring attractor models (Skaggs et al., 1994; Zhang, 1996; Redish et al., 1996; Xie et al., 2002), which typically define a core set of two to three cell types to implement local excitation, global inhibition, and bump movement, raising open questions about the role and necessity of each cell type.

For our experiments, we use connectomic data from the Drosophila FlyEM hemibrain dataset (Scheffer et al., 2020b) after a simple preprocessing step that takes advantage of the natural symmetry between the left and right brain hemispheres to correct synapse counts through a symmetrization process (Section A.1). We focused our analysis on a subpopulation of neurons previously implicated in head direction (HD) processing, comprised of a population of 439 neurons spanning six cell types: E-PG, Delta7, P-EG, P-EN, GLNO, and ring neurons. Notably, the ring neuron population, which is responsible for conveying sensory inputs to the circuit, are divided by lineage into ER and ExR subtypes, which can further be divided into a diverse set of 29 different subtypes of ring neurons.

The connectome provides signed synapse counts (where signs are inferred based on measured neurotransmitters (Davis et al., 2020)) but lack key parameters such as synaptic gains or time constants, and include asymmetries that likely result from biological and measurement noise. This motivates the need for methods that can infer minimal functional parameters from structure while remaining faithful to the measured connectivity.

## 3 Related Work

Connectome-based models have been used to study neural circuit function across a range of scales and assumptions. Chang et al. (2023) tested whether the required global inhibition in the fly head-direction (HD) circuit is provided by Delta7 or ring neurons, using spiking models hand-constructed from connectome-derived motifs and evaluating bump-based metrics across manual parameter sweeps. Stentiford et al. (2024) built a detailed spiking model of the central complex using prespecified physiological parameters, connectivity, and receptive fields, and showed that the models could learn visual cue-to-heading associations via Hebbian plasticity. Pospisil et al. (2024) inferred a whole-brain "effectome" from optogenetic perturbation data, using the connectome as a structural prior along with full-brain activity recordings to infer the effective weights of a linear model. Other works (Mi et al., 2023; Lu et al., 2025) have leveraged neural recordings to infer connectivity structure. Our work is most similar to Lappalainen et al. (2024), who train a deep network constrained by the fly visual motion connectome. However, they assume an idealized, perfectly tiled columnar architecture, and optimize cell-type parameters via supervised learning on an optic flow task. In contrast, we employ

 https://doi.org/10.52202/085713-4273

unsupervised training directly on the raw, measured connectome. Finally, though not a connectome-constrained model, Schaeffer et al. (2023) also used self-supervised objectives to train a circuit to learn an internal representation of a spatial variable, and introduce a conformal isometry loss similar in concept to our linear consistency loss to enforce proportional updates of neural representations with input magnitude. However, whereas they train a randomly initialized network with activity-level constraints, we instead initialize a network with measured connectomic structure, and train it to learn biologically consistent representations through cell-type-level parameterization and self-supervised loss terms without explicit neural activity constraints.

Our approach differs in several key ways. We operate directly on the raw connectome without architectural simplification, activity recordings, or external supervision. We introduce a self-supervised learning objective based on internal representational consistency, which enforces that changes in internal velocity reflect changes in input velocity without requiring activity measurements or task labels, and optimize only a small set of biologically interpretable, cell-type–level parameters. Our method recovers continuous attractor dynamics despite known physiological asymmetries and irregularities of the connectome. Because the objective encodes generic dynamical constraints rather than task-specific labels, our framework can be extended to other circuits hypothesized to implement a coarse computational function, such as integration, memory, or normalization, offering a scalable approach to functional inference from connectomic structure without activity recordings.

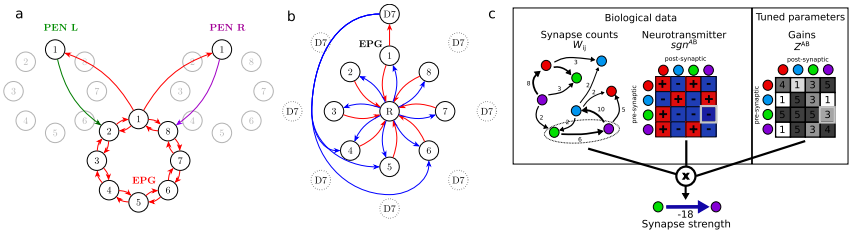



Figure 1: **The HD circuit and model setup. a.** The connectivity pattern of the PEN and EPG neurons. The left/right PEN neurons asymmetrically project to the EPG neurons, causing the ring to rotate clockwise or counterclockwise depending on net input drive. **b.** Ring neurons provide uniform inhibition to EPG neurons, while Delta7 neurons target EPG neurons far away on the ring. **c.** Schematic of tuning procedure. Fixed synapse counts and signs inferred from neurotransmitters are combined with trained cell-type-specific gain parameters to produce the final network weights. Nodes are colored by cell type.

# 4 Methods

## 4.1 Model description

**Rate-based neuron dynamics model** We modeled the dynamics of this neural circuit using a rate-based dynamics model. The firing rate $x_j$ of each neuron $j$ in the circuit evolves according to the following differential equation:

$$\tau \frac{dx_j(t)}{dt} = -\ell\, x_j(t) + \sigma\left(\sum_k W_{jk} x_k(t) + b_j + u_j(t)\right) \tag{1}$$

where $x_j(t)$ is the time-varying firing rate of neuron $j$, $\tau_j$ represents a global intrinsic time constant of each neuron and $\ell$ is a global constant determining the scale of activity and, together with $\tau$, the rate of activity decay in the absence of input. $\sigma(\cdot)$ denotes the sigmoid function, $W_{jk}$ represents the synaptic weight from neuron $k$ to neuron $j$. $b_j$ is the threshold parameter for neuron $j$, and $u_j(t)$ represents any (potentially time-varying) external inputs injected into neuron $j$.

**Connectomics-based model parameterization** To model conectomically defined circuits, we define the weights in Eq. 1 as $W_{ij} = w_0(1+Z_{ij})sgn_{ij}C_{ij}$ where the connectivity matrix ($C$) and the signs ($sgn \in \{\pm 1\}$) of the interaction are fully determined by the empirical synapse counts measured in the connectome and from recorded neurotransmitters, respectively (Scheffer et al., 2020b). The

value $w_0$ is a common gain factor that relates synapse count to connection strength; $Z_{ij}$ corresponds to an additional, unknown differential synapse-specific gain; changes in this parameter permit deviations from connectomically-defined structure. The connectome does not specify neural time constants, neural thresholds, and the neural non-linearity (transfer function). Thus, the parameters $\{Z, b, \tau\}$ are the subjects of optimization.

To examine the extent to which the connectome alone provides sufficient information for detailed dynamics, we constrained the optimization problem with biological inductive biases by assuming that the gain and bias parameters $Z, b$ are shared across all neurons of a given cell type. This parameterization at the level of cell types is biologically motivated, as neurons of the same type exhibit similar gene expression, intrinsic electrophysiological properties, and connectivity patterns (Davie et al., 2018; Turner-Evans et al., 2020). Thus, instead of optimizing parameters $Z_{jk}, b_j$ for each for each synapse and neuron, we optimized a much smaller set of cell-type-specific parameters $Z^{AB}, b^A$, where $A, B, \ldots \in \mathcal{C}$ index cell type. In other words, $Z^{AB}$ is the shared gain factor for the all synapses from neurons of Type B to those of Type A and $b^A$ is the threshold of all neurons of type $A$. The dynamics are therefore given by:

$$\tau \frac{dx_j^A(t)}{dt} = -\ell\, x_j^A(t) + \sigma\left(\sum_{B\in\mathcal{C}}\sum_k w_0(1+Z^{AB})sgn^B C_{jk} x_k^B(t) + b^A + u_j(t)\right) \tag{2}$$

This reduces the number of optimized parameters from $439^2+439+1 = 193,161$ to just $7^2+7+1 = 57$ parameters, relative to a full per-synapse parameterization.

### 4.2 Task-based optimization

**Velocity integration task** We optimize the free network parameters over a set of simulation trials using a set of biologically-motivated self-supervised and unsupervised objectives under the assumption that the circuit performs integration of its input signals.

**State initialization** For each simulation trial, we initialized the firing rates of a single E-PG neuron to 1, representing a bump on the ring, and initialized the activities of all other neurons to 0.

Dynamics were simulated according to Eq. 2 for a fixed duration of $T_{sim} = 2s$ using the Dormand-Prince method (Dopri5) (Dormand and Prince, 1980) provided in the Diffrax (Kidger, 2021) library in Jax. During each trial, a constant velocity input $u \in [-U, U]$ representing clockwise (negative), counter-clockwise (positive), or zero velocity) was applied to the network. Following experimental evidence that GLNO neurons provide angular velocity signals to the HD system (Hulse et al., 2023), the velocity input was selectively injected into either the left or right GLNO neurons.

**Training Objectives** We define a set of biologically motivated objectives based on the assumption that the network is an integrator, i.e. that it should update its internal state as a linear function of the velocity input, and maintain its existing state in the absence of inputs. Notably, we do not specify particular activity profiles in the network, nor do we ask the state to change by a prespecified amount in response to specific velocity inputs.

In each trial, we evaluate the network state at discrete time points $t = (1, \ldots, T)\Delta t$ separated by interval $\Delta t$. We use population coding (Bialek et al., 1989) to decode the network state (location along the ring, $\theta(t)$) at each $t$ from the EPG neuron population by calculating the heading as the vector sum of neurons' preferred directions weighted by their firing rates. The change in the represented angle over the interval $\Delta t$ was computed as $\delta\theta(t) = \frac{\theta(t)-\theta(t-1)}{\Delta t}$.

- **Entropy**: Encourages diverse activation across the population by penalizing overly concentrated or uniform activity patterns.

$$\mathcal{L}_{\text{entropy}} = -\frac{1}{T}\sum_t^T \sum_j^N \phi_j(t)\log\phi_j(t)$$

  Where $\phi_j(t) = \frac{e^{x_j(t)}}{\sum_k^N e^{x_k(t)}}$ is the softmax of the neural activities at a given time point.

- **Stability**: Encourages stability, i.e., no changes in state given zero input ($u = 0$).

$$\mathcal{L}_{\text{stability}} = \frac{1}{T}\sum_t^T \delta\theta(t)^2 \cdot \mathbf{1}(u=0)$$

 https://doi.org/10.52202/085713-4273

- **Minimum Speed**: Prevents the model from learning a trivial zero solution.

$$\mathcal{L}_{\text{speed}} = -\frac{1}{T}\frac{1}{2U}\sum_{t}^{T}\sum_{u=-U}^{U}\delta\theta(t)^2\Theta(|u|)$$

where $\Theta$ is the heavy-side step function.

- **Linear Consistency**: Encourages the model to update its internal state as a linear function of the velocity input

$$\mathcal{L}_{\text{linearity}} = \frac{1}{T}\frac{1}{2U}\sum_{t}^{T}\sum_{u=-U}^{U}\left(\frac{\delta\theta(t)}{|u|} - \mu\right)^2 \Theta(|u|)$$

where $\mu = \frac{1}{T}\frac{1}{2U}\sum_{t}^{T}\sum_{u=-U}^{U}\frac{\delta\theta(t)}{|u|}$

- **L1 and L2 Regularization**: Imposes costs for changes in weights changes, with the penalty size corresponding to the number of synapses that are impacted by each weight.

$$\mathcal{L}_{\text{L1}} = \sum_{i,j\,:\,i\in A,\,j\in B}|Z^{AB}|C_{ij}$$

$$\mathcal{L}_{\text{L2}} = \sum_{i,j\,:\,i\in A,\,j\in B}\left(Z^{AB}C_{ij}\right)^2$$

- **Total loss:**

$$\mathcal{L}_{\text{total}} = \beta_1\mathcal{L}_{\text{linearity}} + \beta_2\mathcal{L}_{\text{speed}} + \beta_3\mathcal{L}_{\text{entropy}} + \beta_4\mathcal{L}_{\text{L1}} + \beta_5\mathcal{L}_{\text{L2}} \tag{3}$$

Where $\beta_1, \beta_2, \beta_3, \beta_4, \beta_5$ are hyperparameters.

**Optimization** The initial values of $Z^{AB}$ are all set to 0. Initial values of $b^A$ are set to a constant $b$ for all cell types. The quantities $w_0, b, \ell$ were chosen by hyperparameter optimization. All trainable parameters were optimized to minimize the total loss function $\mathcal{L}_{\text{total}}$ using Adam (Kingma, 2014).

## 5 Results

### 5.1 Robust integrator dynamics from minimal training of a connectome-constrained model

We optimize model parameters according to the procedure outlined above (Figure 2a). Allowing only a single gain factor ($w_0$) does not yield activity bumps; allowing two global gain factors (one for all excitatory synapses and another for all inhibitory ones) yields an activity bump but no integration. By tuning the small set of cell-type parameters defined above, we recover a dynamical system that is able to integrate bidirectional velocity inputs as well as maintain a stable heading representation in the presence of noise, displaying near-zero drift under constantly injected multiplicative noise at or below 1% of neuron activity (Figure 2b, c, d). Despite not specifying for output shape in our training, we find that the model produces an activity bump with a width (Figure 2e) close to previous experimental reports of $\pi/2$ (Kim et al., 2017). The model also accurately integrates velocity input, maintaining internal state updates that scale proportionally with input magnitude, and minimal pinning region (no bump movement to non-zero velocity inputs), enabling accurate integration over small velocity inputs and velocity direction changes (Figure 2f, g).

### 5.2 Tuned network exhibits characteristics of a ring attractor network

To evaluate whether the trained model exhibits dynamics characteristic of a continuous ring attractor, we analyzed the geometry of its internal state space. Principal component analysis of EPG activity during continuous, time-varying velocity inputs in both directions reveals that the network's population dynamics are constrained to a one-dimensional manifold (Figure 3a, b). This observation is confirmed by local intrinsic dimensionality analysis using local PCA (Kambhatla and Leen, 1997), which estimates the dimensionality of the dynamics to be approximately 1 (Figure 3c). The correlation dimension Grassberger and Procaccia (1983), which quantifies the global scaling properties of the attractor, is also close to 1 (Figure 3d), consistent with a smooth, low-dimensional integrator.

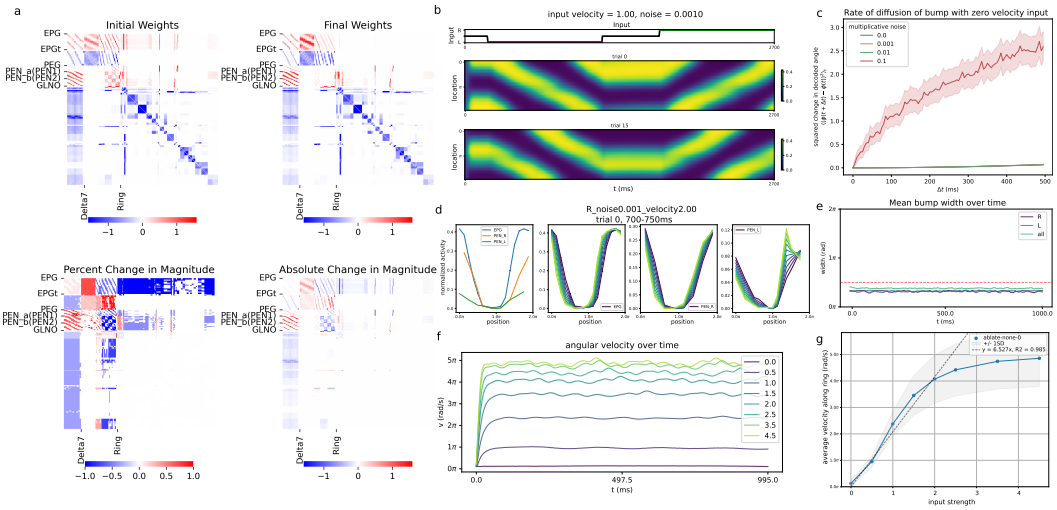



Figure 2: **Tuned network weights produce robust integrator dynamics. a.** Initial and trained network weights. **b.** The activity bump of the EPG neurons across no input and directional input conditions. **c.** The bump exhibits minimal drift for noise levels below 10% of neuron activity **d.** Slices of population activity for EPG and left/right PEN neurons, colored by time. Asymmetric activation of PEN populations drives bump movement in the corresponding direction. **e.** The width of the activity bump, when calculated over left, right, and all neurons, closely matches experimental measurements of $\pi/2$. **f.** The internal heading velocity increases proportionally to the magnitude of input velocity, indicating accurate integration. **g.** The input-output velocity relationship remains linear across a biologically plausible input range.

To further assess the topological structure of the dynamics, we apply persistent homology to the E-PG state trajectories using the Ripser package (Bauer, 2021; Tralie et al., 2018). This analysis yields a persistence diagram consistent with a 1D topological ring (Figure 3e), matching theoretical expectations for a circular continuous attractor. Together, these analyses demonstrate that the network does not merely perform velocity integration in the output space, but that its internal dynamics form a structured, low-dimensional ring manifold, in line with theoretical models of the *Drosophila* HD circuit.

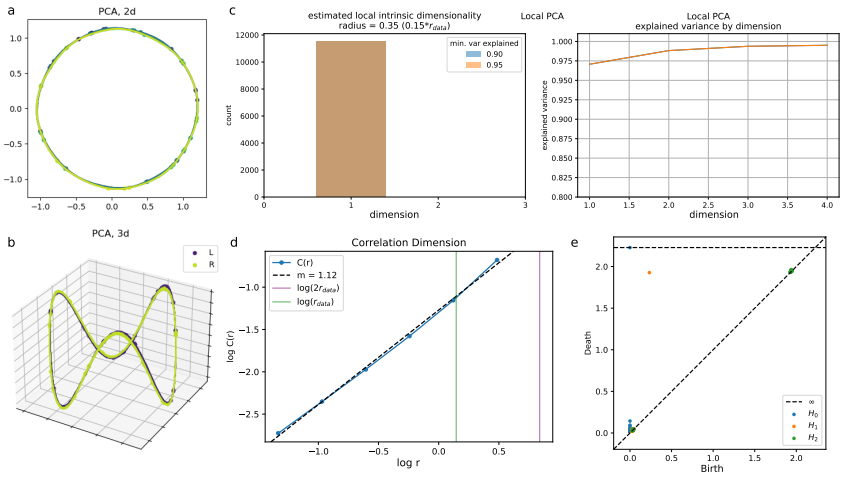



Figure 3: **State space geometry of the HD circuit a.** 2D projections of EPG activity via PCA show that the dynamics lie on a one-dimensional ring manifold. **b.** Ring manifold structure is preserved in a 3D projection in PCA space. **c.** Local PCA estimates the intrinsic dimensionality of the dynamics to be 1. **d.** The local correlation dimension of the network is also approximately 1. **e.** Persistent homology analysis yields a persistence diagram consistent with a 1D topological ring.


https://doi.org/10.52202/085713-4273

### 5.3 The connectome as a structural prior for functional dynamics

Given that finetuning the HD network by training cell-type parameters is sufficient to obtain a network that produces ring attractor dynamics despite noise in the connectome, we hypothesized that the connectome provides a structural prior that supports robust integrator dynamics. To test this, we trained models initialized with increasingly perturbed versions of the connectome. We found that small amounts of Gaussian noise, when preserving connection signs, still yielded integrating solutions. However, introducing sign flips degraded performance, and introducing more drastic perturbations, including shuffling weights within cell-type blocks as well as across the entire connectome, consistently failed to produce functional networks even after extensive training (Figure 4a).

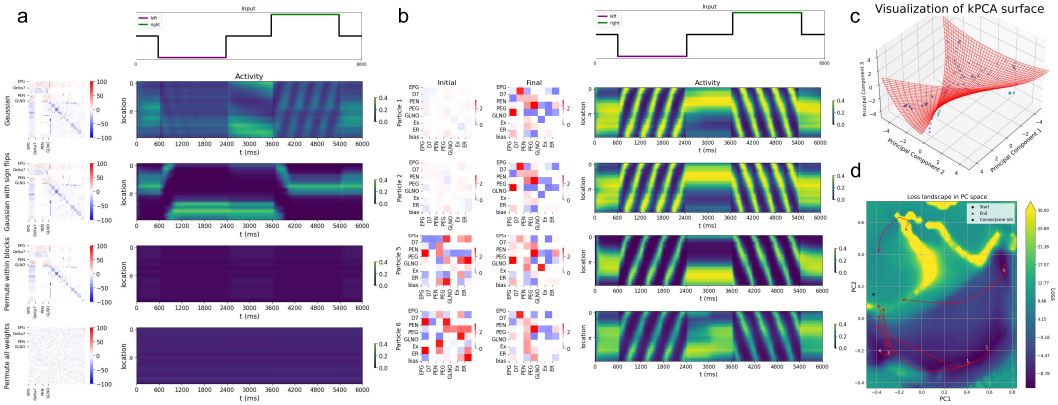



Figure 4: **Visualizing the space of solutions a.** From top to bottom: initialization weights and activity bumps of networks initialized with connectomes perturbed with gaussian noise ($\sigma^2 = 10$), gaussian noise with random sign flips ($p = 0.05$), permutation of weights within cell-type blocks, and permutation of all weights. **b.** Training runs with multiple initial weight matrix conditions. Left: Initial and learned cell-type weights ($Z^{AB}$) centered at 1.0. Right: Heatmap depicting the neural activity over time when left and right inputs are given. **c.** Visualization of the surface fit by kernel PCA. The training trajectory points are used to fit the surface and are depicted by the blue dots. **d.** Visualization of training trajectories along the loss landscape. Trajectories are labeled similarly to b for comparison.

**Cell-type parameterization as the correct level of abstraction** Although the connectome provides essential structural information, learning cell-type-specific parameters is critical for producing a functional network. We trained networks using simplified parameterizations by either tying parameters globally across all cell types, or by learning only broad excitatory and inhibitory cell type parameters (see Section A.2, Figure A4). Neither approach yielded a network capable of velocity integration. Overall, our results showed that a network capable of angular velocity integration does not emerge generically from any connectome, nor from arbitrary choices of biophysical parameters.

**Diversity of solutions from initial positions on the loss landscape** We next explored the degeneracy of the loss landscape by adding different amounts of noise to the network parameters at initialization, and tracking their trajectories across training (Figure 4b). We visualize these trajectories on the fitted loss landscape (Section A.3) of solutions using kernel PCA (Schölkopf et al., 1998). We find that the final solutions of the network depend on the proximity of the initial parameters; networks close by tend to converge to solutions within the same basin, while networks initialized far apart tend to arrive at different solutions. Despite this divergence, networks that arrived at different solutions still produced functional integrator dynamics and reside in the same overall loss basin, indicating some degeneracy in the solution space but a common structural motif (Figure 4c, d).

### 5.4 Cell-type ablations

Given a functional model of the *Drosophila* head direction circuit, we next asked: what are the unique roles of each of the cell types within the circuit? To address this question, we performed ablation

experiments by zeroing the output weights of specific cell-type populations in our model. Because the model weights were optimized on the full set of cell types, we retrained the ablated networks to give each model a fair chance at recovering performance by altering the weights of the remaining cell types. Thus, we test which sets of neurons are strictly necessary in the sense that their function cannot be substituted by simply tuning the weights of other cell types within the network. As with the optimization of the original model, we sweep over a range of hyperparameters and select a set of the best performing models based on linear integration consistency and maximum bump velocity for evaluation (Section A.4). Across all experiments, ablated models exhibit less stability of the bump in the presence of large amounts of noise compared to the original network even after training (Figure 5).

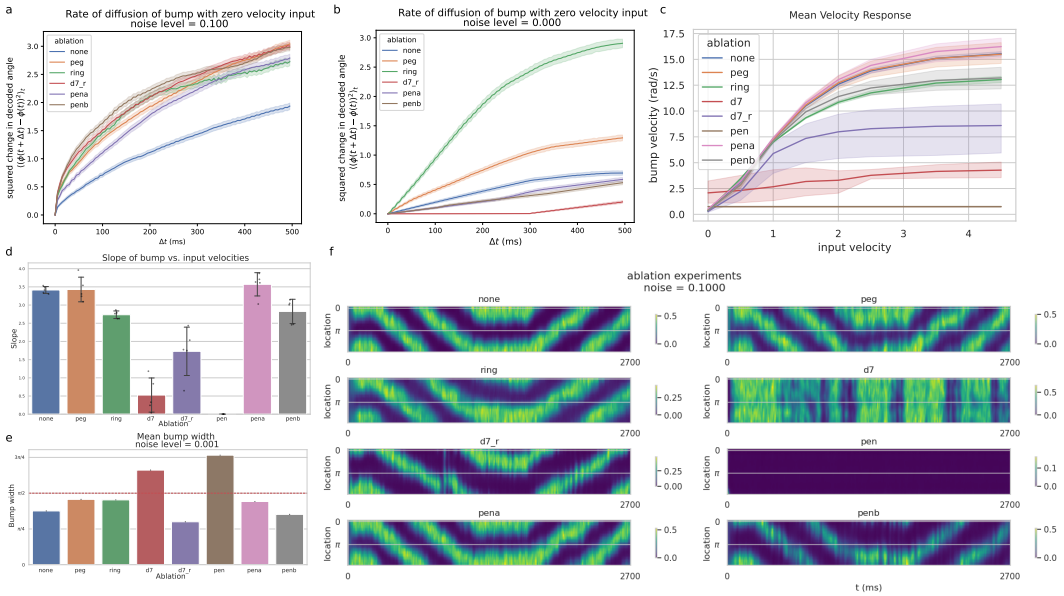



Figure 5: **Cell-type ablation experiments**. **a.** Rate of diffusion of the bump with zero inputs in the presence of multiplicative noise at 10% of neuron activity. Non-integrating solutions (Delta7 and P-EN ablations) not shown. **b.** Rate of diffusion in a noiseless condition. **c.** Mean bump velocity as a function of input velocity, colored by ablation. **d.** Slope of mean velocity response, calculated over inputs in the range of [0, 2]. **e.** Mean bump width of different ablation conditions. Red line indicates experimentally observed width. **f.** Sample EPG activity profiles by ablation, given the same input profile as in Figure 2.

Both Delta7 and ring neurons have been hypothesized to provide the crucial global inhibition to implement a ring attractor in the *Drosophila* (Chang et al., 2023). Although we found that ablating the ring neuron population reduced the range of represented velocities along the ring and displayed more drift in the presence of noise, ablating the Delta7 neurons had a far stronger effect: the network was no longer able to form a bump, let alone integrate velocity (Figure 5f). Notably, we found that the Delta7-ablated network was able to recover bump formation and integration when we relaxed the block matrix assumption and individually tune the weights of all the ring neuron subtypes (Section A.5), indicating that the *Drosophila* HD circuit relies on multiple types of inhibition, in contrast to theoretical ring attractor models.

The P-EG neurons have previously been hypothesized to play a role in stabilizing the bump in the *Drosophila* HD circuit (Pisokas et al., 2020), although experimental evidence in support of this is lacking. We find support for this hypothesis, as the P-EG-ablated model produces a network that exhibits more drift under even low amounts of noise even after training (Figure 5a, b).

Finally, although the role of the P-EN neurons in providing asymmetric drive to move the bump in the E-PG neuron population has been well characterized, the roles of the P-ENa and P-ENb subpopulations remain poorly understood. We confirm the role of the P-EN neurons by showing that a PEN-ablated network is unable to maintain or move a bump even after training (Figure 5f). We then separately ablated the P-ENa and P-ENb neuron subpopulations and found that while ablating


https://doi.org/10.52202/085713-4273

either subpopulation increased drift under noise, the P-ENb ablation had a stronger effect, leading to greater drift and a reduced range of bump velocities.

## 6 Conclusion and Discussion

We present a method that leverages biological priors and parameterization at the level of cell types to recover circuit dynamics from noisy connectome measurements using self-supervised learning. Without access to biophysical parameters, external task labels, or activity data, our model is able to recover robust continuous attractor dynamics capable of stable velocity integration in spite of observed asymmetries and small network size in the *Drosophila* head direction circuit. Interestingly, this result implies that connectomes, without biophysical gain parameters but with knowledge of connectivity signs, may be close to being lottery tickets (Frankle and Carbin, 2018). We characterize the solution space and show that multiple viable solutions emerge across different initializations, allowing estimation of parameter uncertainty through noise injection. We performed in silico experiments via targeted cell-type-specific ablations to elucidate the mechanistic roles of different cell types within the circuit.

Despite being able to recover circuit dynamics, the fidelity of the inferred parameters in this study to the underlying biological parameters may be limited in part by the fidelity of the connectome, because we performed additional (albeit minor) pre-processing of the connectivity matrix and we found a multiplicity of possible solutions. In addition, we made assumptions about neurotransmitter identity and synaptic signs based on cell types.

We believe our self-supervised learning approach will be powerful for future work on inferring function and activity from connectomes, relative to supervised approaches which require prior knowledge about circuit activity states. However, our stability and linear consistency losses terms assumed equivariance to inputs and involved population decoding, limiting their direct application to circuits that integrate continuous variables, such as oculomotor integrators (Seung, 1996), grid cells (Burak and Fiete, 2009), time cells (Kraus et al., 2013), or circuits involved in evidence accumulation (Shadlen and Newsome, 2001). However, we believe our framework could be easily extended by modifying the combination of losses used during training. An important future direction is to generalize our approach by exploring alternative unsupervised objectives that relax these assumptions. Additional future directions include extending the model to study multi-modal sensory integration with the internal state, and validating model predictions through closed-loop perturbation experiments.

## References


Ulrich Bauer. Ripser: efficient computation of Vietoris-Rips persistence barcodes. *J. Appl. Comput. Topol.*, 5(3):391–423, 2021. ISSN 2367-1726. doi: 10.1007/s41468-021-00071-5. URL `https://doi.org/10.1007/s41468-021-00071-5`.

William Bialek, Fred Rieke, Robert van Steveninck, and David Warland. Reading a neural code. *Advances in neural information processing systems*, 2, 1989.

Yoram Burak and Ila R Fiete. Accurate path integration in continuous attractor network models of grid cells. *PLoS computational biology*, 5(2):e1000291, 2009.

Ning Chang, Hsuan-Pei Huang, and Chung-Chuan Lo. Global inhibition in head-direction neural circuits: a systematic comparison between connectome-based spiking neural circuit models. *Journal of Comparative Physiology A*, 209(4):721–735, 2023.

Steven J Cook, Travis A Jarrell, Christopher A Brittin, Yi Wang, Adam E Bloniarz, Maksim A Yakovlev, Ken CQ Nguyen, Leo T-H Tang, Emily A Bayer, Janet S Duerr, et al. Whole-animal connectomes of both caenorhabditis elegans sexes. *Nature*, 571(7763):63–71, 2019.

Kristofer Davie, Jasper Janssens, Duygu Koldere, Maxime De Waegeneer, Uli Pech, Łukasz Kreft, Sara Aibar, Samira Makhzami, Valerie Christiaens, Carmen Bravo González-Blas, et al. A single-cell transcriptome atlas of the aging drosophila brain. *Cell*, 174(4):982–998, 2018.

Fred P Davis, Aljoscha Nern, Serge Picard, Michael B Reiser, Gerald M Rubin, Sean R Eddy, and Gilbert L Henry. A genetic, genomic, and computational resource for exploring neural circuit function. *Elife*, 9:e50901, 2020.

Sven Dorkenwald, Arie Matsliah, Amy R Sterling, Philipp Schlegel, Szi-Chieh Yu, Claire E McKellar, Albert Lin, Marta Costa, Katharina Eichler, Yijie Yin, et al. Neuronal wiring diagram of an adult brain. *Nature*, 634(8032):124–138, 2024.

John R Dormand and Peter J Prince. A family of embedded runge-kutta formulae. *Journal of computational and applied mathematics*, 6(1):19–26, 1980.

Jonathan Frankle and Michael Carbin. The lottery ticket hypothesis: Finding sparse, trainable neural networks. *arXiv preprint arXiv:1803.03635*, 2018.

Peter Grassberger and Itamar Procaccia. Measuring the strangeness of strange attractors. *Physica D: nonlinear phenomena*, 9(1-2):189–208, 1983.

Jonathan Green, Atsuko Adachi, Kunal K Shah, Jonathan D Hirokawa, Pablo S Magani, and Gaby Maimon. A neural circuit architecture for angular integration in drosophila. *Nature*, 546(7656): 101–106, 2017.

Brad K Hulse, Angel Stanoev, Daniel B Turner-Evans, Johannes D Seelig, and Vivek Jayaraman. A rotational velocity estimate constructed through visuomotor competition updates the fly's neural compass. *bioRxiv*, page 2023.09.25.559373, September 2023.

Nandakishore Kambhatla and Todd K Leen. Dimension reduction by local principal component analysis. *Neural computation*, 9(7):1493–1516, 1997.

Mikail Khona and Ila R Fiete. Attractor and integrator networks in the brain. *Nature Reviews Neuroscience*, 23(12):744–766, 2022.

Patrick Kidger. *On Neural Differential Equations*. PhD thesis, University of Oxford, 2021.

Sung Soo Kim, Hervé Rouault, Shaul Druckmann, and Vivek Jayaraman. Ring attractor dynamics in the drosophila central brain. *Science*, 356(6340):849–853, 2017.

Diederik P Kingma. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

Benjamin J Kraus, Robert J Robinson, John A White, Howard Eichenbaum, and Michael E Hasselmo. Hippocampal "time cells": time versus path integration. *Neuron*, 78(6):1090–1101, 2013.

Janne K Lappalainen, Fabian D Tschopp, Sridhama Prakhya, Mason McGill, Aljoscha Nern, Kazunori Shinomiya, Shin-ya Takemura, Eyal Gruntman, Jakob H Macke, and Srinivas C Turaga. Connectome-constrained networks predict neural activity across the fly visual system. *Nature*, 634 (8036):1132–1140, 2024.

Ziyu Lu, Wuwei Zhang, Trung Le, Hao Wang, Uygar Sümbül, Eric Todd SheaBrown, and Lu Mi. Netformer: An interpretable model for recovering dynamical connectivity in neuronal population dynamics. In *The Thirteenth International Conference on Learning Representations*, 2025.

Lu Mi, Trung Le, Tianxing He, Eli Shlizerman, and Uygar Sümbül. Learning time-invariant representations for individual neurons from population dynamics. *Advances in Neural Information Processing Systems*, 36:46007–46026, 2023.

MICrONS Consortium. Functional connectomics spanning multiple areas of mouse visual cortex. *Nature*, 640:435–447, 2025. doi: 10.1038/s41586-025-08790-w. URL `https://www.nature.com/articles/s41586-025-08790-w`.

Ioannis Pisokas, Stanley Heinze, and Barbara Webb. The head direction circuit of two insect species. *Elife*, 9:e53985, 2020.

Dean A Pospisil, Max J Aragon, Sven Dorkenwald, Arie Matsliah, Amy R Sterling, Philipp Schlegel, Szi-chieh Yu, Claire E McKellar, Marta Costa, Katharina Eichler, et al. The fly connectome reveals a path to the effectome. *Nature*, 634(8032):201–209, 2024.

A David Redish, Adam N Elga, and David S Touretzky. A coupled attractor model of the rodent head direction system. *Network: computation in neural systems*, 7(4):671, 1996.

Rylan Schaeffer, Mikail Khona, Tzuhsuan Ma, Cristobal Eyzaguirre, Sanmi Koyejo, and Ila Fiete. Self-supervised learning of representations for space generates multi-modular grid cells. *Advances in Neural Information Processing Systems*, 36:23140–23157, 2023.

Louis K Scheffer, C Shan Xu, Michal Januszewski, Zhiyuan Lu, Shin-Ya Takemura, Kenneth J Hayworth, Gary B Huang, Kazunori Shinomiya, Jeremy Maitlin-Shepard, Stuart Berg, Jody Clements, Philip M Hubbard, William T Katz, Lowell Umayam, Ting Zhao, David Ackerman, Tim Blakely, John Bogovic, Tom Dolafi, Dagmar Kainmueller, Takashi Kawase, Khaled A Khairy, Laramie Leavitt, Peter H Li, Larry Lindsey, Nicole Neubarth, Donald J Olbris, Hideo Otsuna, Eric T Trautman, Masayoshi Ito, Alexander S Bates, Jens Goldammer, Tanya Wolff, Robert Svirskas, Philipp Schlegel, Erika Neace, Christopher J Knecht, Chelsea X Alvarado, Dennis A Bailey, Samantha Ballinger, Jolanta A Borycz, Brandon S Canino, Natasha Cheatham, Michael Cook, Marisa Dreher, Octave Duclos, Bryon Eubanks, Kelli Fairbanks, Samantha Finley, Nora Forknall, Audrey Francis, Gary Patrick Hopkins, Emily M Joyce, Sungjin Kim, Nicole A Kirk, Julie Kovalyak, Shirley A Lauchie, Alanna Lohff, Charli Maldonado, Emily A Manley, Sari McLin, Caroline Mooney, Miatta Ndama, Omotara Ogundeyi, Nneoma Okeoma, Christopher Ordish, Nicholas Padilla, Christopher M Patrick, Tyler Paterson, Elliott E Phillips, Emily M Phillips, Neha Rampally, Caitlin Ribeiro, Madelaine K Robertson, Jon Thomson Rymer, Sean M Ryan, Megan Sammons, Anne K Scott, Ashley L Scott, Aya Shinomiya, Claire Smith, Kelsey Smith, Natalie L Smith, Margaret A Sobeski, Alia Suleiman, Jackie Swift, Satoko Takemura, Iris Talebi, Dorota Tarnogorska, Emily Tenshaw, Temour Tokhi, John J Walsh, Tansy Yang, Jane Anne Horne, Feng Li, Ruchi Parekh, Patricia K Rivlin, Vivek Jayaraman, Marta Costa, Gregory Sxe Jefferis, Kei Ito, Stephan Saalfeld, Reed George, Ian A Meinertzhagen, Gerald M Rubin, Harald F Hess, Viren Jain, and Stephen M Plaza. A connectome and analysis of the adult drosophila central brain. *Elife*, 9, September 2020a.

Louis K Scheffer, C Shan Xu, Michal Januszewski, Zhiyuan Lu, Shin-ya Takemura, Kenneth J Hayworth, Gary B Huang, Kazunori Shinomiya, Jeremy Maitlin-Shepard, Stuart Berg, et al. A connectome and analysis of the adult drosophila central brain. *elife*, 9:e57443, 2020b.

Bernhard Schölkopf, Alexander Smola, and Klaus-Robert Müller. Nonlinear component analysis as a kernel eigenvalue problem. *Neural computation*, 10(5):1299–1319, 1998.

Johannes D Seelig and Vivek Jayaraman. Neural dynamics for landmark orientation and angular path integration. *Nature*, 521(7551):186–191, 2015.

H S Seung. How the brain keeps the eyes still. *Proc. Natl. Acad. Sci. U. S. A.*, 93(23):13339–13344, November 1996.

Michael N Shadlen and William T Newsome. Neural basis of a perceptual decision in the parietal cortex (area lip) of the rhesus monkey. *Journal of neurophysiology*, 86(4):1916–1936, 2001.

William Skaggs, James Knierim, Hemant Kudrimoti, and Bruce McNaughton. A model of the neural basis of the rat's sense of direction. *Advances in neural information processing systems*, 7, 1994.

Rachael Stentiford, James C Knight, Thomas Nowotny, Andrew Philippides, and Paul Graham. Estimating orientation in natural scenes: A spiking neural network model of the insect central complex. *PLOS Computational Biology*, 20(8):e1011913, 2024.

Christopher Tralie, Nathaniel Saul, and Rann Bar-On. Ripser.py: A lean persistent homology library for python. *The Journal of Open Source Software*, 3(29):925, Sep 2018. doi: 10.21105/joss.00925. URL `https://doi.org/10.21105/joss.00925`.

Daniel Turner-Evans, Stephanie Wegener, Hervé Rouault, Romain Franconville, Tanya Wolff, Johannes D Seelig, Shaul Druckmann, and Vivek Jayaraman. Angular velocity integration in a fly heading circuit. *Elife*, 6:e23496, 2017.

Daniel B Turner-Evans, Kristopher T Jensen, Saba Ali, Tyler Paterson, Arlo Sheridan, Robert P Ray, Tanya Wolff, J Scott Lauritzen, Gerald M Rubin, Davi D Bock, et al. The neuroanatomical ultrastructure and function of a biological ring attractor. *Neuron*, 108(1):145–163, 2020.

Xiaohui Xie, Richard HR Hahnloser, and H Sebastian Seung. Double-ring network model of the head-direction system. *Physical Review E*, 66(4):041902, 2002.

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

## 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: The abstract makes summarizes the methods and results that were achieved in the main paper.

   Guidelines:

   - The answer NA 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 NA 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: Limitations have been discussed in the discussion portion of the paper.

   Guidelines:

   - The answer NA 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: [NA]

Justification: This paper provides empirical results about ring attractor networks and does not present new theoretical results.

Guidelines:

- The answer NA 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: The authors provide an anonymized version of the code and data.

   Guidelines:

   - The answer NA 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?


https://doi.org/10.52202/085713-4273

Answer: [Yes]

Justification: The authors provide the data and code.

Guidelines:

- The answer NA means that paper does not include experiments requiring code.
- Please see the NeurIPS code and data submission guidelines (`https://nips.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://nips.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, etc.) necessary to understand the results?

   Answer: [Yes]

   Justification: The authors provide the experimental details in the methods section.

   Guidelines:

   - The answer NA 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: Statistical significance is reported via error bars in the main paper when appropriate.

   Guidelines:

   - The answer NA 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: The authors provide the compute resources used in the Hardware and Software section.

   Guidelines:

   - The answer NA 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: This paper conforms to the NeurIPS Code of Ethics.

   Guidelines:

   - The answer NA 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: [NA]

    Justification: This paper concerns basic research on ring attractor networks and does not pose any direct societal impacts.

    Guidelines:

    - The answer NA means that there is no societal impact of the work performed.
    - If the authors answer NA 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.


https://doi.org/10.52202/085713-4273

- 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., pretrained language models, image generators, or scraped datasets)?

    Answer: [NA]

    Justification: This paper uses publicly available data and the models are specialized for the drosophila head direction system and have no applications outside of this.

    Guidelines:

    - The answer NA 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: The authors provide the licenses for the existing assets in the Hardware and Software License section.

    Guidelines:

    - The answer NA 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: [NA]

    Justification: This paper does not introduce new assets.

    Guidelines:

    - The answer NA 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: [NA]

    Justification: This paper does not involve crowdsourcing or research with human subjects.

    Guidelines:

    - The answer NA 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: [NA]

    Justification: This paper does not involve research with human subjects.

    Guidelines:

    - The answer NA 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.

 https://doi.org/10.52202/085713-4273

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 rigorousness, or originality of the research, declaration is not required.

Answer: [NA]

Justification: This paper does not use LLMs as an important, original, or non-standard component of the core methods.

Guidelines:

- The answer NA 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 (`https://neurips.cc/Conferences/2025/LLM`) for what should or should not be described.

# A Technical Appendices and Supplementary Material

## A.1 Weight Symmetrization

To enforce bilateral symmetry in the neural network architecture while preserving functional connectivity patterns, we implemented a weight matrix symmetrization procedure. Let $\mathbf{W} \in \mathbb{R}^{N \times N}$ be the initial weight matrix, where $N$ is the number of neurons. For each neuron $i$, we define its lateralization $l_i \in \{L, R\}$ and its lateralized label $\tilde{l}_i$ (removing lateralization information).

The symmetrization process operates on connection groups defined by the tuple $(s, t, l_s, l_t)$, where $s$ and $t$ are the lateralized labels of the source and target neurons, and $l_s$ and $l_t$ are their lateralizations. For each unique connection group, we compute the symmetrized weight:

$$w_{sym}(s, t, l_s, l_t) = \frac{1}{2n_{st}} \sum_{i,j} w_{ij} \mathbb{I}[(s, t, l_s, l_t) = (\tilde{l}_i, \tilde{l}_j, l_i, l_j)]$$

where $n_{st}$ is the number of connections in the group and $\mathbb{I}$ is the indicator function. The final symmetrized weight matrix $\mathbf{W}_{sym}$ is constructed as:

$$W_{sym,ij} = \begin{cases} w_{sym}(\tilde{l}_i, \tilde{l}_j, l_i, l_j) & \text{if } \exists w_{sym}(\tilde{l}_i, \tilde{l}_j, l_i, l_j) \\ W_{ij} & \text{otherwise} \\ 0 & \text{if } i = j \end{cases}$$

This procedure ensures that left-right symmetric connections have identical weights ($W_{sym,ij} = W_{sym,ji}$ for lateralized pairs), while preserving the network's functional connectivity patterns and maintaining the overall network structure and strength. The result of this symmetrization is shown in Figure A1.

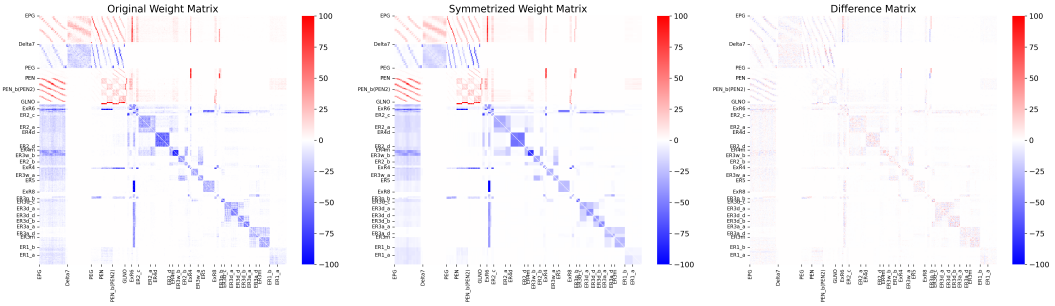



Figure A1: Weight matrix symmetrization.

All experiments we're run with weight symmetrization but eliminating weight symmetrization is possible if left and right populations of each cell type are treated as separate cell populations which can be tuned independently. This doubles the number of cell types but comparable results are achievable using this parameterization, even when weight symmetrization is not applied A2.

## A.2 Single parameter and E/I controls

In addition to the cell-type parameterization of the model, we also performed baseline experiments using a single global parameter to parameterize the synapse gains across all cell-types (thus not tuning differential gains between cell types), as well as using two parameters to describe the excitatory and inhibitory neuron gains separately. These reduced parameterizations failed to produce solutions capable of forming or integrating a bump even after model training, demonstrating the importance of cell-type parameterization (Figure A4).

## A.3 Kernel PCA and loss landscape visualization

In order to visualize the optimization landscape, we ran training runs with differently seeded initial parameters. We tracked the parameter values throughout training. We then used kernel PCA


https://doi.org/10.52202/085713-4273

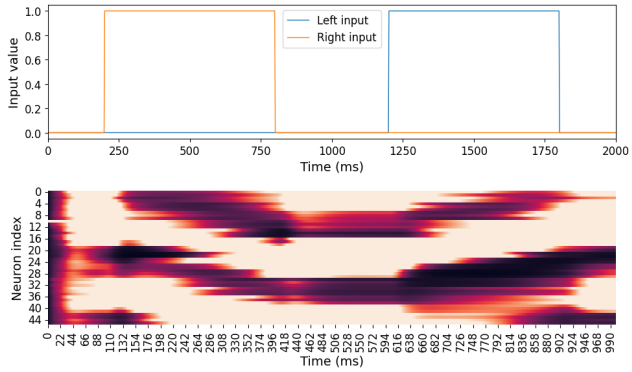



Figure A2: Neural dynamics after weight symmetrization is removed but left and right hemispheres are allowed to have different biophysical parameterizations.

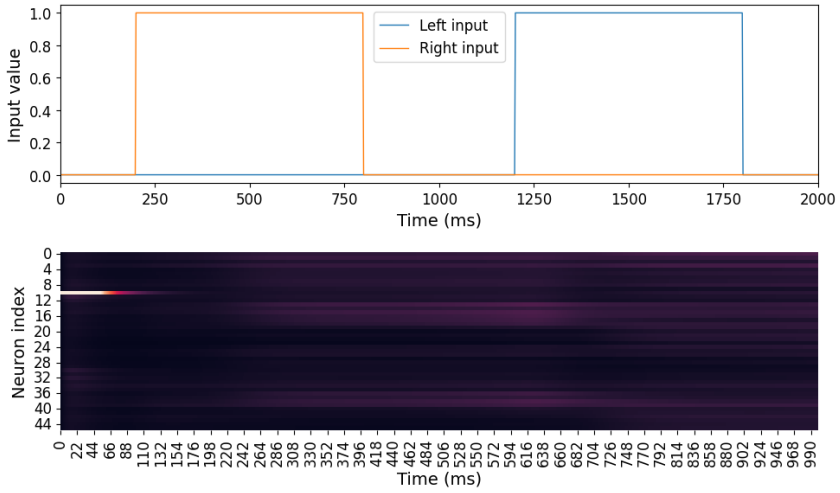



Figure A3: Visualization of the dynamics of an untrained network.

(Schölkopf et al., 1998) with a radial basis function kernel to fit the training trajectories and used a Kernel Ridge regression to learn the mapping from the basis to the original parameter space which we use to sample parameter values on in order to construct a two dimensional loss landscape.

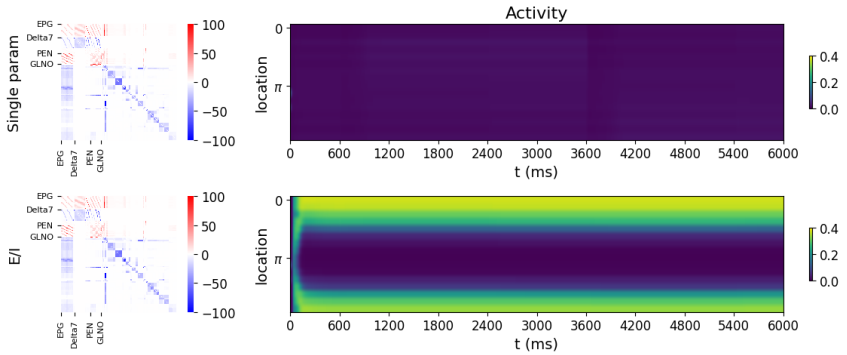



Figure A4: Baseline experiments. Learned weights and ring activity after optimizing only a single global gain parameter $w_0$ (top), or optimizing two parameters describing excitatory and inhibitory synaptic gains (bottom).

### A.4 Hyperparameter sweeps and model selection

For each ablation experiment (including the no-ablation condition), we trained models across a set of 90 hyperparameters by using a grid search over the initial bias values, leak, and global synapse strength. We additionally randomly initialized the block weights using a normal distribution centered at 1 with variance of 2 percent of the parameter value and swept over five seeds for each hyperparameter setting. In Figure 5 we chose the best 5 models from each condition to evaluate. We define best as the models with the smallest minimum speed loss (i.e., the models with the greatest range of velocity integration), after filtering for a R2 of > 0.7 and a temporal consistency loss of < 0.013.

### A.5 Delta7 ablations with full ring parameterization

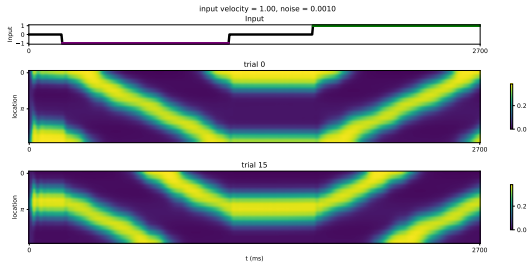



Figure A5: Activity bumps in Delta7 ablation, full ring parameterization.

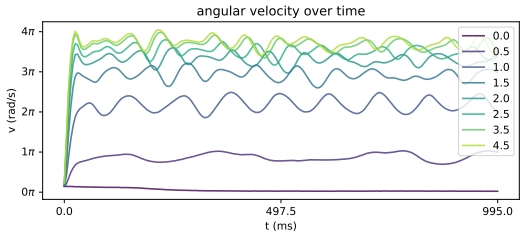
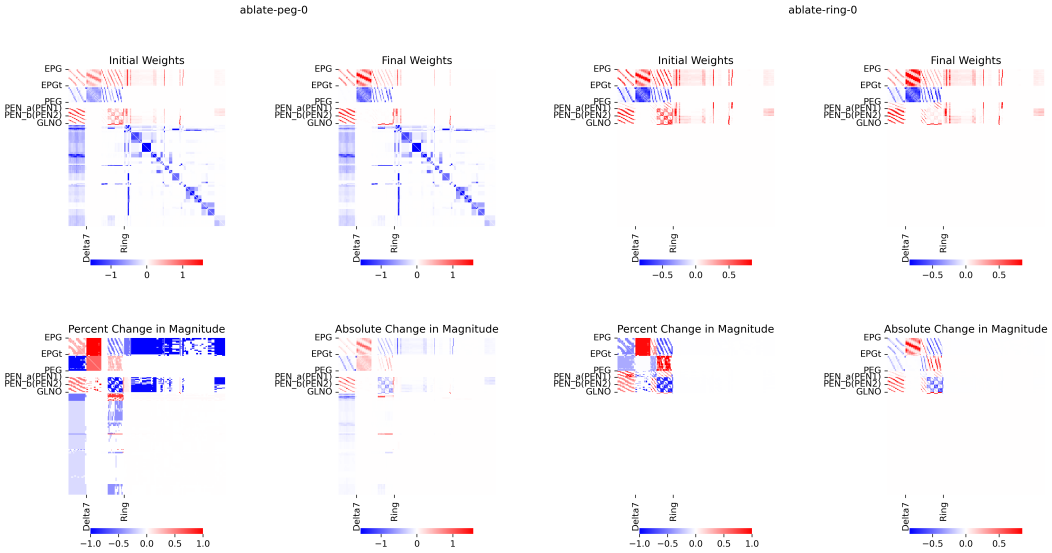



Figure A6: Internal vs. input velocity in Delta7 ablation, full ring parameterization.

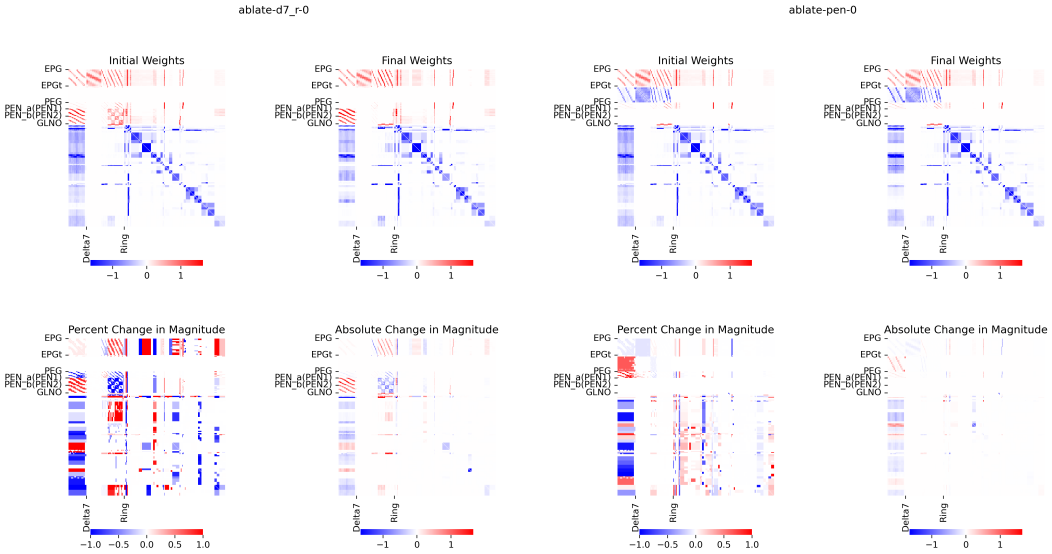

https://doi.org/10.52202/085713-4273

## A.6 Sample weights from ablation experiments




(a) P-EG ablation

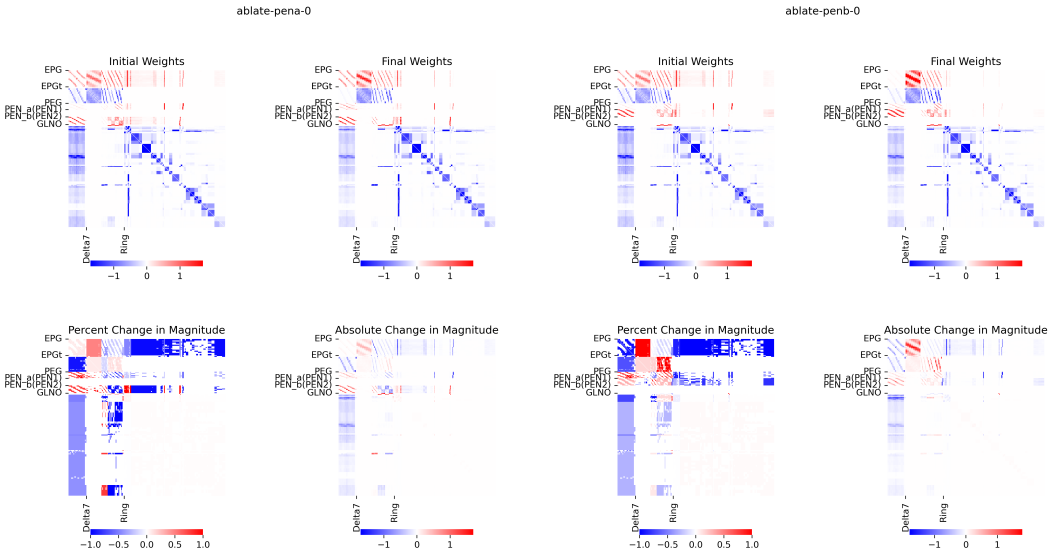
(b) Ring ablation




(c) Delta7, full ring parameterization

(d) P-EN ablation

Figure A7: Ablation effects on trained synaptic weights: P-EG, Ring, Delta7, and P-EN.




(a) P-ENa ablation (b) P-ENb ablation

Figure A8: Ablation effects on trained synaptic weights: P-ENa and P-ENb.

## A.7 Additional Supplementary Methods

**Velocity input** Velocity input is driven by differential activity between the left and right GLNO neurons. In our experiments, we drove the velocity using either the left or right GLNO neurons at a given time. It is sufficient to drive differential activity between the left and right GLNO neurons to drive the velocity input which can be shown in A10.


https://doi.org/10.52202/085713-4273

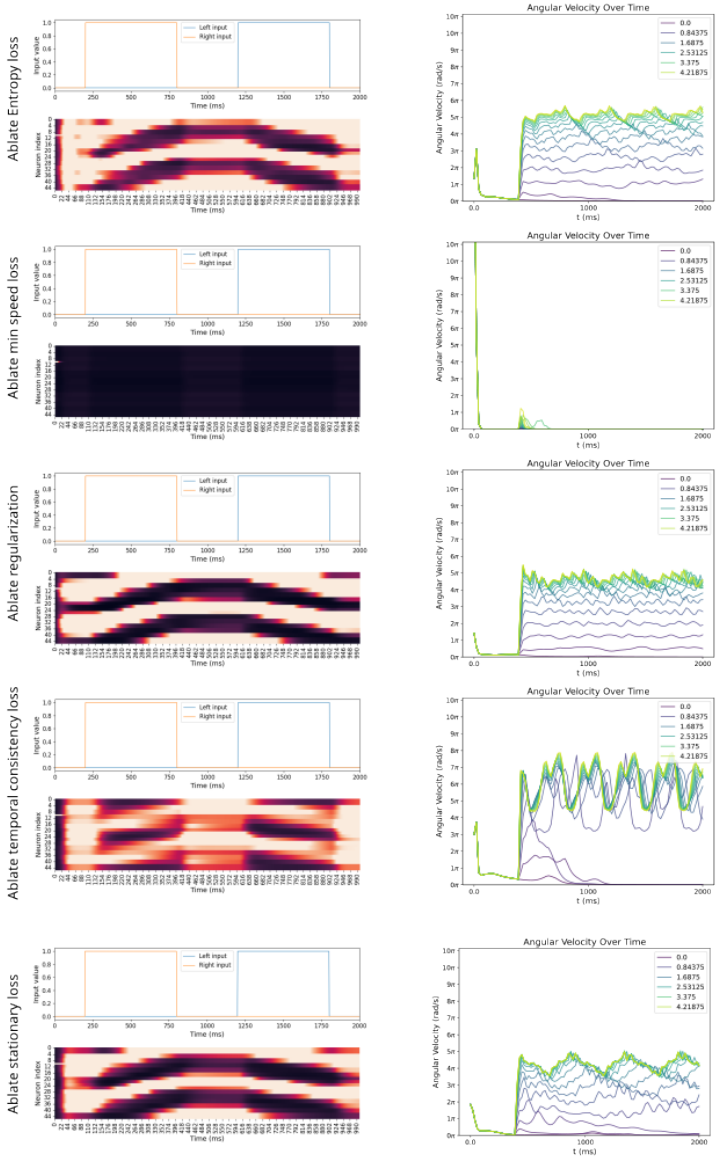



Figure A9: Neural Dynamics after ablating various loss terms. We show an example neural trajectory on the left hand side rotating in both directions as well as the angular velocity over time under different input conditions

https://doi.org/10.52202/085713-4273

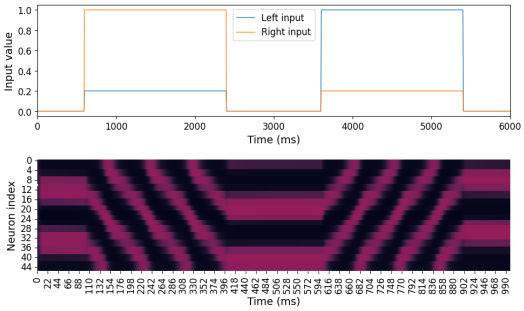



Figure A10: Activity trace when reversing sign of velocity input

## A.8 Supplementary Figures

 https://doi.org/10.52202/085713-4273

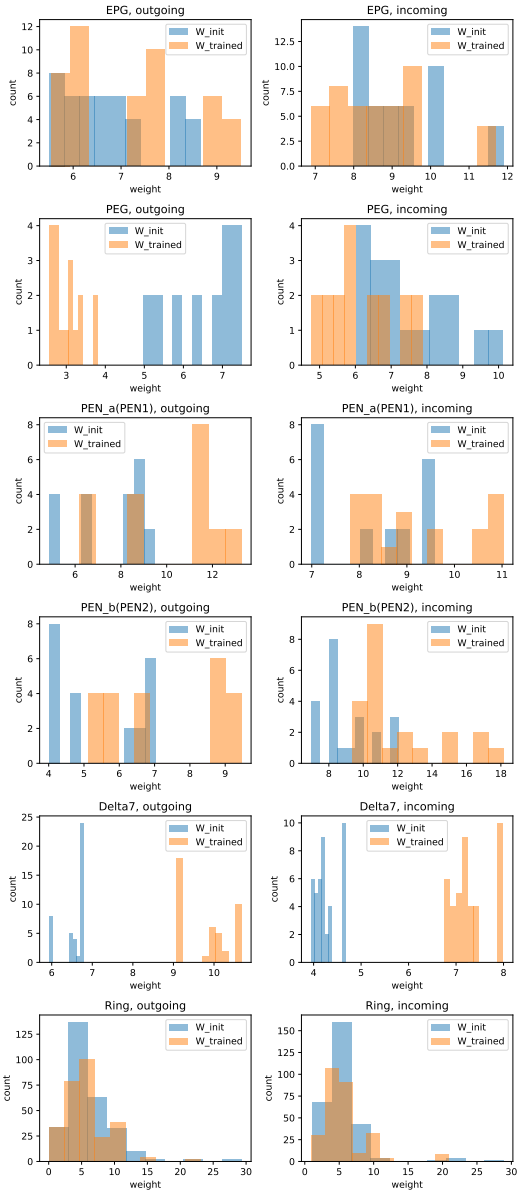



Figure A11: **Synapse statistics.** Distribution of incoming and outgoing synapse weights before and after training of a full connectome-initialized network.