Predicting Neural Activity from Connectome Embedding Spaces

Zihan Zhang[1], **Huanqiu Zhang**[2], **and Stefan Mihalas**[1]

[1]Center for Data-Driven Discovery in Biology, Allen Institute, Seattle, WA 98109, USA
[2]Data Science Institute & Department of Neurobiology, The University of Chicago, Chicago, IL 60637, USA

Describing the relationship between activity and connectivity in neural circuits is fundamental to understanding brain function. The connectome of cortical networks is high-dimensional, whereas neuronal activity is often low-dimensional relative to the number of neurons. Consequently, only a small fraction of the information contained in the connectome is relevant to activity. Can this relevant information be identified? In this study, we propose that the activity of each excitatory neuron in the sensory cortex can be approximated as a point in an embedding space defined by its connectome. Using the MICrONS dataset, which provides millimeter-scale, nanometer-resolution data on proofread and coregistered neurons, we perform a correlation analysis and identify a statistically significant alignment between morphological and functional similarities. Topological analysis reveals that the structured representation spaces of both neuronal activity and connectome share a low-dimensional, hyperbolic geometry with exponential scalability. Based on these findings, we develop a simple linear model to reconstruct in vivo neuronal activity by embedding anatomical affinities using multidimensional scaling (MDS). This approach achieves an explanation ratio of 68% compared to directly using activity similarity, outperforming the 56% obtained using the full connectome. Our results highlight a clear and robust structure-function coupling: although geometry-compatible dimensionality reduction methods discard much of the connectome's detailed information, they prove more effective in predicting neuronal activity, suggesting that synaptic connections may encode an abstract low-dimensional organization.

1: Introduction

The human brain is a complex multiscale structure, encompassing various molecular, cellular, and neuronal phenomena that constitute the foundation of cognition (1, 2). Experimental evidence estimates that the average human brain contains over 86 billion neurons and approximately 10^{14} synaptic connections (3, 4), while the mouse brain comprises approximately 85 million neurons, forming roughly 100 billion synaptic connections (5). Each synapse, depending on its strength, has been suggested to store up to 4.7 bits of information (6). However, are all these structural details essential to define the circuit activity in the brain? For the first time, based on monumental efforts (7), we now have the capability to compare simplified models that link structure to activity in cortical circuits under varying levels of coarse-graining in connectivity. While detailed structural properties shape brain activity, the extent to which they are necessary for functional modeling remains an open question. Simple artificial neural networks (ANNs) without bio-realistic constraints may not serve as ideal models, as their material and metabolic limitations, synaptic signals, and anatomical scaffolds are not encoded in the same manner as parameters in ANNs. Consequently, the mechanics and functionalities of the brain may not be fully explained or replicated by the scaling laws employed in artificial systems (8–13). Recent studies have found that, despite the complex and high-dimensional nature of individual neuronal responses, population activity across neurons often unfolds with low-dimensional subspaces, particularly in sensory and motor systems (14–17). Although these studies suggest that, due to its low-dimensional nature, activity contains only a smaller amount of information than neurons can independently encode, no link to its emergence from structure is proposed.

Historically, the nontrivial relationship between structural connectivity ($\mathbf{W}$) and functional characteristics ($\mathbf{A}$) of biological neuronal networks has been widely recognized, as exemplified by classical Hebbian theory and the three-factor learning rule (18, 19). However, most studies exploring this relationship remain theoretical and lack systematic validation with empirical data. Experimental efforts to establish a direct link between structure and function primarily rely on MRI data, which emphasize macroscopic network-level analyses rather than detailed synaptic-level investigations (20–25). More recently, studies on *C. elegans* and fruit flies (*Drosophila Melanogaster*) have provided detailed connectomic data, enabling finer-grained analyses (26–30). Building on this, (31) demonstrates that the drosophila connectome can serve as a prior to estimate a causal model of the central nervous system, revealing that dominant circuits involve a small neuronal population and enabling efficient whole-brain analysis with sparse imaging and perturbations. Extending this idea, (32) effectively predicts individual neuronal selectivity to dynamic visual stimuli by leveraging parameter sharing within the same cell type. While these connectome-based studies employ state-of-the-art deep learning models to uncover complex high-dimensional relationships between structure and function, our work instead focuses on identifying normative, low-dimensional representations of these relationships. Specifically, we adopt a low-dimensional embedding approach based on vectorized connectome similarity, which has been previously used to predict and explain edgewise functional connectivity in inter-hemispheric homologous regions (33, 34). From a broader perspective, our study aligns with previous work by advocating for the incorporation of biological details into computational models in a simplified

DRAFT

yet principled manner that enhances both performance and metabolic efficiency (31, 32). However, unlike prior studies that emphasize the precise details of the connectome, our approach focuses on a lossy compression of connectomic information, capturing essential structural properties while abstracting away fine-grained complexities.

2: Structure-Function Alignment

We use the largest functionally-imaged electron microscopy (EM) dataset currently available, MICrONS, which was generated through dense calcium imaging followed by the reconstruction of a millimeter-scale volume (7, 35, 36). This dataset provides nanometer-resolution details on the spatial locations of each cell's soma and associated synapses, along with classification labels, postsynaptic density (PSD) sizes, and activity recordings of neurons in the primary visual cortex (V1) as well as the anterolateral (AL), lateromedial (LM), and rostrolateral (RL) higher visual areas. The activity data consists of detrended, deconvolved, and stimulus- and behavior-aligned recordings from excitatory neurons spanning layers 2 to 6. Figure 1A-B provides an illustration of neuronal connectivity within the dataset. To ensure anatomical completeness, we focus on neurons that have undergone manual segmentation and full proofreading of both axons and dendrites (7, 35, 37). Our analysis incorporates results from 12 separate scans, with each scan containing between 35 and 80 neurons. The same inclusion and preprocessing strategies as in prior work are applied to ensure consistency (7, 36). Additional details about the MICrONS dataset used in this study are provided in Section A.

We employ a "correlation of correlation" approach to quantify the relationship between neuronal input/output connection similarity ($\mathbf{C}^{\text{in}}$/$\mathbf{C}^{\text{out}}$) and functional similarity ($\mathbf{C}^{\text{act}}$). The matrices $\mathbf{C}^{\text{in}}$ and $\mathbf{C}^{\text{out}}$ are derived from the non-binary connectome by counting the total number of synapses between pairs of neurons. To assess the global relational structure of the network, we focus on non-local and higher order connectivity measures rather than direct (Hebbian) connection, and therefore do not weight the pairwise relationships by synaptic counts or strengths (38–41). Specifically, $\mathbf{C}^{\text{out}}$ is derived from neurons with fully reconstructed dendrites, while $\mathbf{C}^{\text{in}}$ is based on neurons with extended or clean axons (7). Functional similarity, $\mathbf{C}^{\text{act}}$, is computed from the full neuronal response across each scan, comprising $30,000$ data points and yielding a median pairwise similarity of 1.18% (Figure 1D). Formally, the connectivity matrices are defined as $\mathbf{W}^{\text{in}} \in \mathbb{R}^{M_1 \times N}$ and $\mathbf{W}^{\text{out}} \in \mathbb{R}^{N \times M_2}$, where N denotes the number of targeted neurons (or features), and M_1 and M_2 represent the number of structurally confident neurons for axons and dendrites, respectively, where $M_1, M_2 \gg N$. Section A provides additional statistical details and parameter descriptions. The connection similarity matrix is then computed as:

$$\mathbf{C}^{\text{in}}_{ij} = \text{corr}(\mathbf{W}^{\text{in}}_{:,i}, \mathbf{W}^{\text{in}}_{:,j}),\ \mathbf{C}^{\text{out}}_{ij} = \text{corr}(\mathbf{W}^{\text{out}}_{i,:}, \mathbf{W}^{\text{out}}_{j,:}). \quad \textbf{(1)}$$

$\forall\ 1 \leq i, j \leq N$. As a result, $\mathbf{C}^{\text{in}}, \mathbf{C}^{\text{act}} \in \mathbb{R}^{N \times N}$. We then define the connectivity-activity similarity matrices:

$$\mathbf{C}^{\text{in,act}}_{ij} = \text{corr}\left(\mathbf{C}^{\text{in}}_{i,:}, \mathbf{C}^{\text{act}}_{j,:}\right).\ \mathbf{C}^{\text{out,act}}_{ij} = \text{corr}\left(\mathbf{C}^{\text{in}}_{i,:}, \mathbf{C}^{\text{act}}_{j,:}\right). \quad \textbf{(2)}$$

Namely, each entry in $\mathbf{C}^{\text{in,act}}$ represents the Pearson correlation between the ith row of $\mathbf{C}^{\text{in}}$ and the jth row of $\mathbf{C}^{\text{act}}$. This matrix has diagonal elements representing the correlation between connection and activity similarities for each neuron, while the off-diagonal elements serve as controls, relating the functional similarity of one neuron to the connection similarity of another. The distribution of connectivity-activity similarities for individual neurons is significantly higher than the shuffled control (Figure 1E1). Notably, input similarity provides a more accurate estimate of functional similarity than output similarity (Figure 4). This finding is further supported by Central Kernel Alignment (Figure 1E2) and by analyzing the principal angles between subspaces in the functional and structural affinity spaces (Figure 5) (42, 43).

3: Hyperbolic Geometry for Both Connectome and Activity Representation Spaces

We investigate the underlying geometry of representation space by analyzing neuronal structure and activity using topological data analysis. Specifically, we focus on Betti curves, which quantify cycles (i.e., closed loops in a topological space) and cliques (i.e., fully connected subgraphs in a simplicial complex) in graphs generated by varying the threshold that defines significant connections between neurons (44). Similar topological methods based on persistent homology has been widely applied to study the manifold structures of population neural or perceptual responses in systems such as olfaction, the hippocampus, head direction, and grid cells, consistently revealing that the observed geometry is often hyperbolic rather than Euclidean (45–49).

Specifically, we randomly sample points from the 2D hyperbolic space $\mathbb{H}^2$, using inverse pairwise distance as the correlation measure, and compute the corresponding Betti curves. The similarity between the Betti curves of the synthetic and experimental data serves as a signature of geometric alignment, which remains invariant under nonlinear monotonic transformations of the correlation values (46, 49). This analysis reveals that both connectivity and neuronal dynamics exhibit geometric properties more consistent with low-dimensional hyperbolic space than with Euclidian space, as determined by comparing the mean squared error and the integrated Betti values (Figure 1F). The advantages of hyperbolic geometry become more pronounced as the dataset size increases, as it inherently offers a higher embedding capacity and efficiently accommodate additional nodes (Figure 17, Section G and Section I). This observation remains robust across different connectivity patterns of $\mathbf{W}$ and various correlation measures. Additional methodological details on these analyses are provided in Section E. Furthermore, we apply the Hyperbolic t-SNE algorithm to embed the structural space of proofread neurons in $\mathbb{H}^2$, further supporting the presence of hyperbolic geometry in neuronal organization (Figure 1C) (50).

These findings underscore the relevance of hyperbolic geometry to neuronal organization and reveal a shared geometric

substrate that links structural and functional spaces. The exponential volume growth and low distortion inherent to hyperbolic space enable compact and efficient representations, particularly well-suited for modeling hierarchical, tree-like networks commonly observed in biological systems (51, 52). This structural efficiency not only reduces storage and computational costs but also supports scalable and interpretable embeddings of complex neuronal data. Furthermore, hyperbolic geometry facilitates robust communication protocols that require only local network information, maintaining efficiency even under dynamic changes in connectivity (53, 54). Notably, low-dimensional hyperbolic embeddings capture the underlying structure of the data with comparable fidelity to Euclidean embeddings of much higher dimensionality, thereby offering a principled approach for analyzing and interpreting large-scale neural circuits (Figure 21) (45).

4: Activity Trace Reconstruction through Linear Model

We implement a simple statistical linear model, previously shown to be effective in fMRI analysis (55). Given an activity trace $\mathbf{A} \in \mathbb{R}^{N\times T}$ with N targeted neurons and T time points (Figure 2), neuronal dynamics are reconstructed using:

$$\tilde{\mathbf{A}} = (\mathbf{C} - I) \cdot \mathbf{A}, \tag{3}$$

where $\mathbf{C} \in \mathbb{R}^{N\times N}$ represents pairwise affinities (i.e. weights of dynamical similarity), and $\tilde{\mathbf{A}} \in \mathbb{R}^{N\times T}$ denotes the reconstructed activity. This model predicts the activity of each neuron by leveraging the trajectories of all other neurons in a leave-one-out manner. To evaluate model performance, we compute the reconstruction accuracy a as the average Pearson correlation between corresponding rows in $\tilde{\mathbf{A}}$ and $\mathbf{A}$, enabling neuron-wise comparison. When $\mathbf{C} = \mathbf{C}^{\text{act}}$, a_{act} reaches 14.1%, averaged across 12 scans. This serves as an upper-bound control in the untrainable linear model, as the model directly uses observed activity correlations. To avoid overfitting, we bin the activity data, compute $\mathbf{C}^{\text{act}}$ using the concatenated information from odd-numbered bins, and evaluate a_{act} on the remaining data. To predict activity from structure without accessing the activity of the targeted neuron, $\mathbf{C}$ can be replaced with either the in-degree ($\mathbf{C}^{\text{in}}$) or out-degree ($\mathbf{C}^{\text{out}}$) correlation matrices derived from the connectome, allowing structural similarity to serve as the predictor. Alternatively, instead of using the full connectomic, samples can be projected into a low-dimensional space using non-metric multidimensional scaling (MDS) to either hyperbolic or Euclidean space, optimizing for stress and rank preservation (50, 56). The inverse distance in the chosen metric space, standardized appropriately, is then used as the pairwise affinity measure. Notably, activity correlation varies substantially across scans due to changes in the mouse's behavioral state during visual stimulus presentation. This results in significant variability in absolute reconstruction accuracy, largely driven by a small subset of neurons (36). To mitigate this effect, we focus on relative performance, quantified as the explanation ratio a/a_{act}, which measures the effectiveness of alternative affinity measures relative to $\mathbf{C}^{\text{act}}$ (20).

In the baseline experiment, glial cells are excluded to focus on interactions between excitatory (E) and inhibitory (I) neurons. Synaptic sign is assigned according to Dale's law, with inhibitory neurons having negative output weights and excitatory neurons having positive output weights. Neuronal identity is determined based on morphological features (7). $\mathbf{C}^{\text{in}}$ and $\mathbf{C}^{\text{out}}$ are computed using the same approach described in Section 2. Synaptic information from all functionally coregistered neurons across scans ($N = 497$), including neurons without complete structural confidence, is concatenated and subjected to dimensionality reduction. Using the input connectome correlation, we observe an explanation ratio of 56.3%, whereas the output connectome yields 35.6% (Figure 3A). While the effectiveness of using correlations derived from the input connectome are intuitive, the ability of the output connectome correlations to predict activity is less immediately clear (18, 57). Previous studies employed a "digital twin" model to predict neural responses to a diverse stimulus set of natural and parametric movies, characterizing the *in silico* functional properties of imaged neurons. These studies revealed an overarching like-to-like principle – neurons with highly correlated responses to natural videos tend to be connected across multiple layers and visual cortical areas – providing insights into the effectiveness of $\mathbf{C}^{\text{out}}$ (36, 58, 59). The observation that $\mathbf{C}^{\text{in}}$ yields better accuracy than $\mathbf{C}^{\text{out}}$ is consistent with the discussion in Section 2. Furthermore, having more neurons included in the scan (larger N) improves reconstruction accuracy when using connectome or activity similarity, suggesting that further experimental refinement may improve performance (Figure 22). However, the explanation ratio remains stable and uncorrelated to N, highlighting the robustness of the observation (Figure 23). Moreover, we evaluate the performance of direct synaptic connectivity – another common approach to quantifying $\boldsymbol{\alpha}$ – within the linear model and find that it performs significantly worse than input similarity, although comparably to output similarity, regardless of the connectivity modality (Figure 33) (20, 60).

Surprisingly, utilizing the embedding results in either geometry improves reconstruction accuracy compared to using the raw connectome correlation (e.g. for hyperbolic embedding, 66.6% for input and 62.0% for output), despite the information per neuron being reduced from N to 2 (Figure 3A). However, using an excessively small or large $R_{\max}$ for hyperbolic embedding degrades performance: a small $R_{\max}$ limits capacity, while a large $R_{\max}$ introduces excessive encoding freedom, increasing stochasticity and requiring more repeated experiments for stability. Encouragingly, the stability range is relatively large (Figure 3B). The explanation ratios for hyperbolic and Euclidean embeddings are statistically similar across scans, likely due to the limited number of coregistered neurons, which reduces structural complexity and introduces a selective bias, resulting in a nonuniform or bimodal distribution caused by nonhomogeneous neuronal identity (see Section G and Figure 9). Noteworthy, the number of neurons that are imaged, reconstructed, and coregistered – 497 in total, with 35-80 neurons per scan – is substantially smaller than the number of neurons that are imaged or reconstructed ($N = 716$). Only the

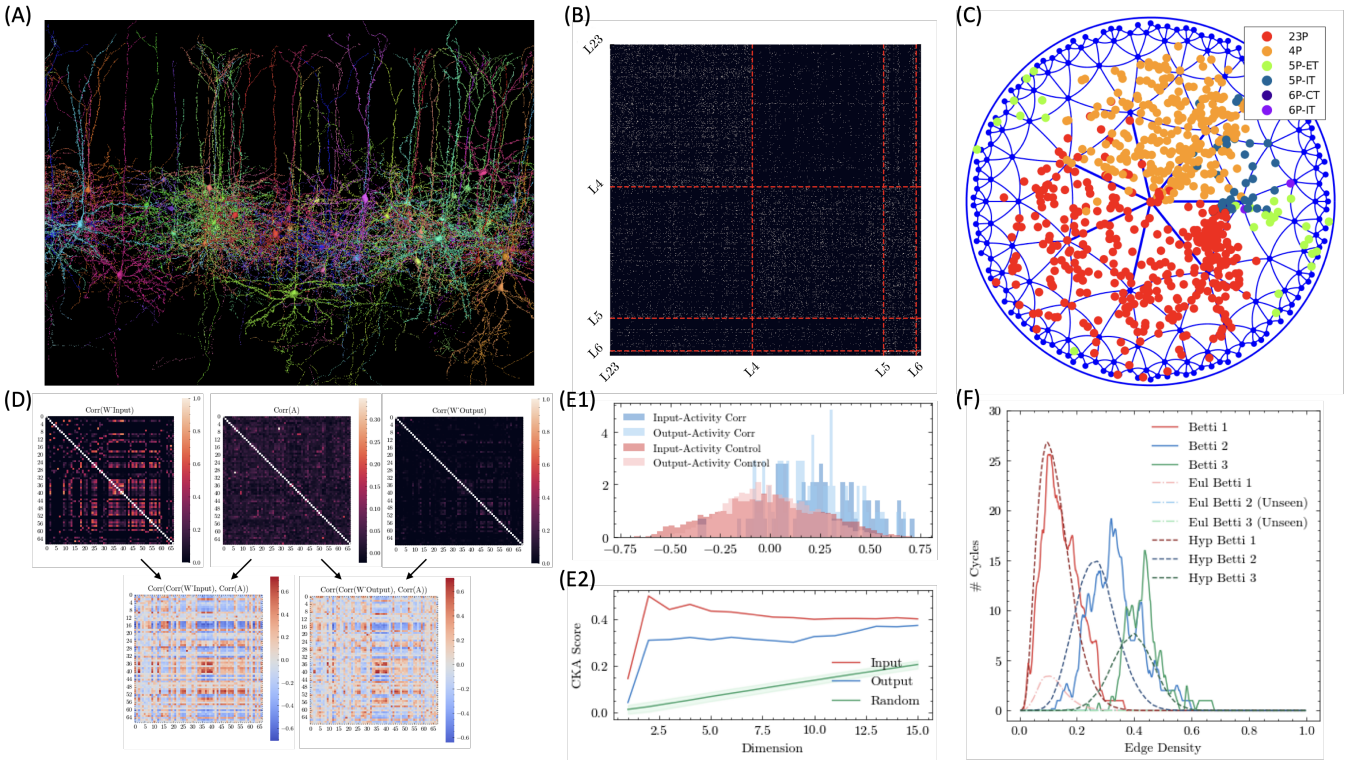

Fig. 1. Demonstration of MICrONS Dataset and Structure-Function Alignment: (A): Illustration of 89 proofread layer 4 pyramidal cells from the MICrONS dataset using `neuroglancer`; (B): Binary connectome showing all-to-all connections between fully proofread excitatory neurons ($N = 716$), ordered by layers. The horizontal axis represents postsynaptic neurons, while the vertical axis represents the presynaptic neurons; (C): Embedding of the structural out-degree distribution of neurons in (B) in $\mathbb{H}^2$ space using the Hyperbolic t-SNE algorithm for clustering by cell type and layer. The projection is shown on a unit circle; (D): Illustration of input connection correlation $\mathbf{C}^{\text{in}}$ (left-top), neuronal activity correlation $\mathbf{C}^{\text{act}}$ (middle-left), output connection correlation $\mathbf{C}^{\text{out}}$ (top-right), alignment between $\mathbf{C}^{\text{act}}$ and $\mathbf{C}^{\text{in}}$ (bottom-left, also see Equation (2)), and alignment between $\mathbf{C}^{\text{act}}$ and $\mathbf{C}^{\text{out}}$ (right-bottom); (E1): Correlation between activity and connection similarity for inputs (dark blue) and outputs (light blue), along their respective controls (dark red and light red). t-test p-values: $p_1 = 9.25 \times 10^{-12}$ (dark blue vs. dark red, testing whether input connections are correlated with activity), $p_2 = 2.24 \times 10^{-14}$ (light blue vs. light red, testing whether output connections are correlated with activity), and $p_3 = 3.66 \times 10^{-1}$ (dark blue vs. light blue, testing whether input connections exhibit a stronger correlation with activity than output connections). See Section B for results from additional scans; (E2): Central Kernel Analysis (CKA) of low-rank approximations using Singular Value Decomposition (SVD) for $\mathbf{C}^{\text{act}}$ with $\mathbf{C}^{\text{in}}$ (red), $\mathbf{C}^{\text{out}}$ (blue), and random matrices (green). The x-axis represents the inclusion of different numbers of top eigenvalues; (F): Betti curves for activity representation spaces show better alignment with points uniformly sampled from 2D Hyperbolic space than from Euclidean space. Standard deviations are omitted for clarity. The averaged results of 100 hyperbolic sampling (dashed line), using the optimized $R_{\max}$ value (see Section E for details), yield a cumulative absolute pointwise error among the top three Betti curves of 14.67. The corresponding Betti values, $\bar{\beta}_m$, are 3.31, 2.46, and 1.21, compared to the groundtruth Betti values of 3.47, 3.08, 1.64. Section H presents a comprehensive Betti analysis of all activity data. Figure (D-F) analyzes Session 8, Scan 5, from the MICrONS dataset (7).

former group is used in the structure-function analysis, while the latter is used for hyperbolic geometry analysis. Detailed statistics are provided in Section A.

Our results remain generalizable and stable when $\mathbf{C}^{\text{in}}$ and $\mathbf{C}^{\text{out}}$ are computed using either the binary connection pattern or postsynaptic density (PSD) volume (Figure 24), or when considering connections exclusively within subpopulations of specific cell types (Figure 25) (61, 62). Compared to the baseline experiment, using the binary connectome yields comparable accuracy in activity reconstruction (56.2% for input and 36.0% for output), whereas incorporating PSD information significantly reduces performance (52.0% and 28.0%, respectively). Furthermore, even partially accurate structural information – specifically, incorporating connections from or to unproofread E/I neurons in the computation of $\mathbf{C}^{\text{in}}/\mathbf{C}^{\text{out}}$ – remains beneficial for activity prediction. The most notable improvement is observed for $\mathbf{C}^{\text{in}}$ (64.6%), with a marginal improvement for $\mathbf{C}^{\text{out}}$ (37.8%) (Figure 26). Embedding-based results remain relatively consistent across different connectivity representations. Additionally, restricting connections from or to certain subpopulations of neurons (e.g. E/I) significantly reduces the dimensionality of $\mathbf{W}$ and limits the performance (36.4% for excitatory input and 48.8% for inhibitory input), whereas the embedding results are less sensitive to this reduction (63.8% for excitatory input and 66.5% for inhibitory input) (Table 1, Figure 25) ((40, 63)).

While we focused on activity throughout the entire experiment, multiple different stimuli were presented to the mice consisting of natural movies, global directional parametric stimuli ("Monet"), and local directional parametric stimuli ("Trippy"), interleaved with multiple off-periods (usually 8-12), and displayed (7, 36). To account for these variations, we compute $\mathbf{C}^{\text{act}}$, reconstruct activity traces separately for each period, and report the averaged results. While explanation ratios decrease due to the increased specificity captured by $\mathbf{C}^{\text{act}}$, the advantage of using embeddings remains evident (see p-values in Figure 27). For completeness, we evaluate our model's performance directly on raw fluorescence data, which exhibits significant population-level pairwise correlations (median of 4.90% and mean of 7.00%). This reduces the significance of input embeddings while preserves that of output embeddings (see p-values in Figure 28 and Table 5).

The performance of embedding results based on input similarity remains comparable to those using $\mathbf{C}^{\text{in}}$, although the statistical significance diminishes. However, the improved performance on the output side is preserved.

Additionally, we explore downsampling by selecting neurons either based on the top-K entries or at random. Specifically, after computing $\boldsymbol{\alpha}$, we filter (zero out) its entries column by column, either by magnitude or randomly, and then predict the leave-one-out neuron using only a subset of the N neurons (Figure 29). Notably, $\mathbf{C}^{\text{act}}$ preserves its accuracy when using only the top-20%, whereas the performance of structurally derived representations deteriorates significantly, particularly for embedded results. For example, in the baseline experiment (Figure 29A), selecting the top 20% entries from $\mathbf{C}^{\text{act}}$ preserves 93.0% of its performance, while only 70.8% and 85.1% are retained for $\mathbf{C}^{\text{in,hyp}}$ and $\mathbf{C}^{\text{in}}$ respectively. The degradation of $\mathbf{C}^{\text{in,hyp}}$ arises because subsampling reduces data quantity, and embedding further compresses dimensionality, together incurring greater information loss compared to the higher resolution and structural density of the original array ($\mathbf{C}^{\text{in}}$). Moreover, the performance advantage of $\mathbf{C}^{\text{in,hyp}}$ over $\mathbf{C}^{\text{in}}$ becomes less pronounced with smaller values of K ($p = 0.003$ for 100% (Figure 3A), 0.051 for 50%, and 0.905 for 20%). However, the significance level is preserved under random selection ($p = 0.001$ for random 50% and 0.002 for random 20%). The observation remains qualitatively consistent across different models (Figure 29).

To further refine our analysis, we restrict the dataset to neurons that are both functionally coregistered and fully proofread ($N = 338$ in total). This filtering enhances reliability by leveraging output similarity in prediction and generally improves relative performance compared to using the unfiltered dataset, partly because the performance of $\mathbf{C}^{\text{act}}$ also declines for small N (Figure 22 and Figure 30). To maintain consistency with the previous geometric analysis (Section G), we perform MDS on all proofread excitatory neurons ($N = 712$) to obtain a universal embedding that preserves relevant interactions across a large subset of neurons. We then extract the corresponding information for the scan-specific coregistered neurons for reconstruction purposes.

So far, we exclusively focus on the 2D embedding. We increase the embedding dimension and surprisingly observed that, despite significant differences in compression quality, the overall reconstruction remains closely aligned, highlighting the model's robustness (Figure 32).

We also construct a null model using the classical continuous-time recurrent neural network (RNN) (64). Without any training, we inject random inputs and compute the correlation of hidden activities in the recurrent layer with 60 neurons. To ensure a fair comparison with the MICrONS dataset, we fine-tune the input dimensions so that the dimensionality of the activity space, quantified through normalized participation ratio, matches that of the experimental data (≈ 0.30) (Figure 31) (65). This selection accounts for the impact of dimensionality on dimensionality reduction methods. To mimic the typical non-negative distribution of deconvolved activity traces (Figure 2), we use ReLU as the activation function, which naturally constrains values in $\mathbf{C}^{\text{act}}$. Notably, projecting random matrices into a low-dimensional space presents empirical challenges, often causing pairwise distances to approach the space limit (Figure 21), leading to a high stress value. Figure 3C illustrates the correlation of the reconstructed results, where $\mathbf{C}^{\text{act}}$ achieves promising performance with $\approx 74\%$ accuracy. In contrast, both $\mathbf{C}^{\text{in}}$ and $\mathbf{C}^{\text{out}}$ follow an approximately Gaussian distribution centered at 0, effectively nullifying the alignment effect through total summation. The weak signal observed in the embedded results can be attributed to the averaged signal from population activity. This comparative study of the random network highlights that biological systems provide a well-structured scaffold with meaningful structure-function alignment, which can be leveraged to infer and adapt information across different levels.

5: Discussion

In this work, we investigate the relationship between synaptic connectivity and the functional responses of excitatory neurons in the mouse visual cortex, leveraging dense reconstruction data from the MICrONS dataset (7, 36, 59). Our analysis reveals a clear alignment between pairwise structural and functional affinities: neurons that receive similar inputs or outputs tend to exhibit similar functional behaviors, and vice versa. Furthermore, by identifying matching topological signatures, we demonstrate that both structural and functional representation spaces can be embedded in low-dimensional hyperbolic spaces with varying radii. Using non-metric MDS method, we construct a correlation matrix based on the spatial proximity of points within these geometric embeddings. Notably, using these results, we achieve not only comparable but significantly improved reconstruction accuracy of activity traces via a simple linear model, despite operating in a significantly lower-dimensional space than the full connectivity matrix. This observation suggests that many anatomical details may not be critical for understanding coarse structure-function mapping and that their interactions can be effectively captured in a reduced parameter space.

It is well established that excitatory neurons with similar responses properties are more likely to be synaptically connected in the mouse primary visual cortex (40, 41, 66–71). Consistent with this, we also observed like-to-like connectivity in our analysis (Figure 6). This phenomenon is proposed to arise in networks where enhanced connectivity strengthens functional covariance and where learning promotes similar responses to similar stimuli (36, 72). Many studies have demonstrated a strong relationship between direct connectivity among targeted neurons ($\mathbf{W} \in \mathbb{R}^{N \times N}$, as opposed to $\mathbf{C}^{\text{in}}$ or $\mathbf{C}^{\text{out}}$) and functional interaction, as shown in human mesoscale projectomes, macaque axonal tract tracing study and mesoscale study in Drosophila (20, 60, 73, 74). More similar to (75), we see that structural similarity is a better predictor of activity than direct synaptic connections, since the consideration of non-local interactions increases the fraction of all potential inputs and drivers, rather than restricting to each neuron's immediate casual inputs (Figure 33). This broader scope could enable the identification of additional structure-function

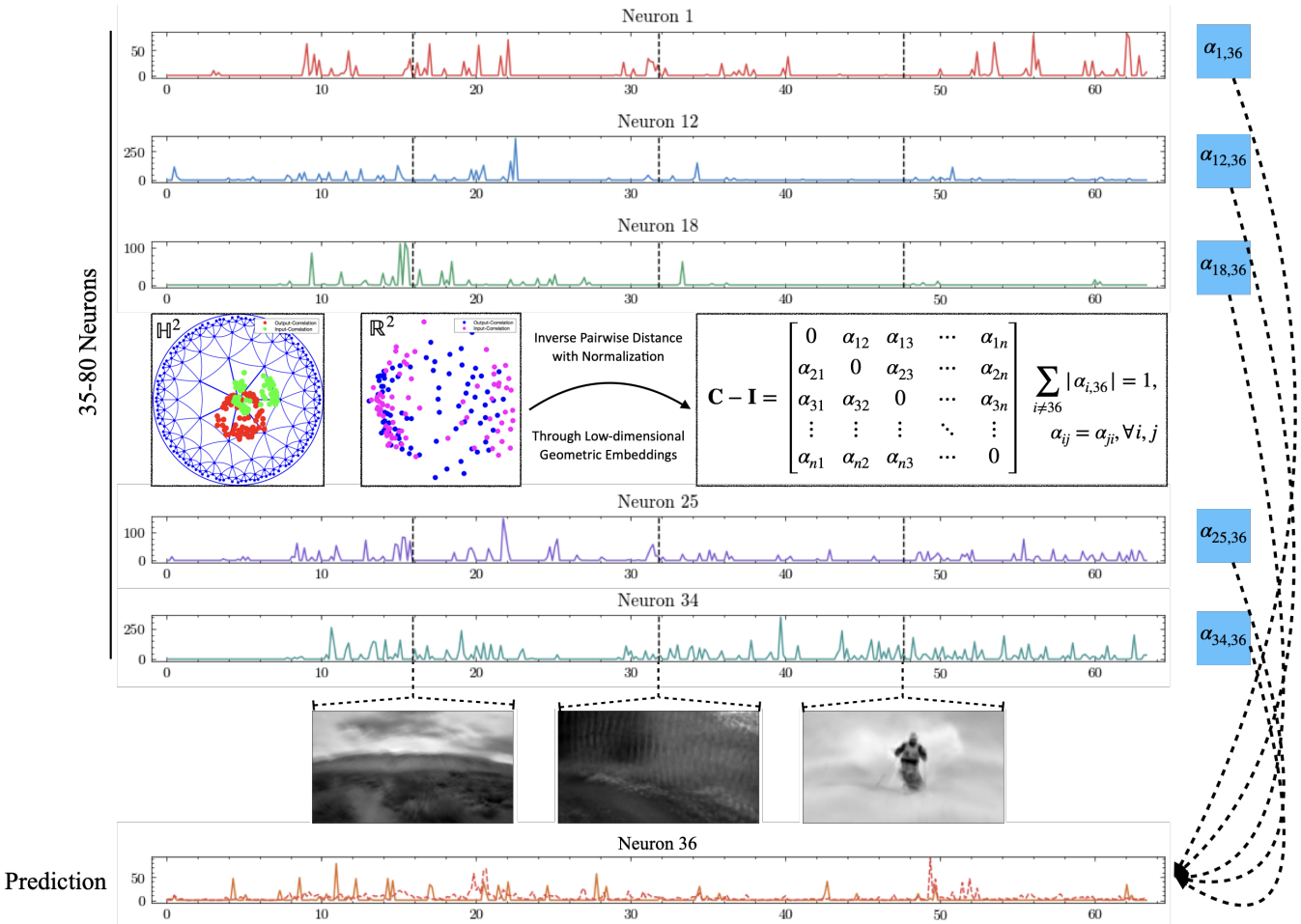

Fig. 2. Schematics of Linear Model: The model utilizes deconvolved calcium traces from coregistered imaged neurons recorded during serial video trails (with sample frames of natural videos shown below) to predict the activity of a held-out neuron (7, 36). The prediction is obtained by computing a weighted sum of activities from other neurons, with weights represented by the symmetric matrix $\mathbf{C}$. Define $\boldsymbol{\alpha} = \mathbf{C} - \mathbf{I}$, where $\boldsymbol{\alpha}_{ij}$ denotes the weight assigned to neuron i's activity trace in predicting the activity of neuron j. We present 1% of the total traces from Session 8, Scan 5 of the MICrONS dataset. Solid line: groundtruth; dashed line: prediction. Frame rate: $6.3/s$. x-axis in seconds. Inset: MDS projections of $\mathbf{C}^{\text{in}}$ and $\mathbf{C}^{\text{out}}$ to low-dimensional Hyperbolic and Euclidean spaces.

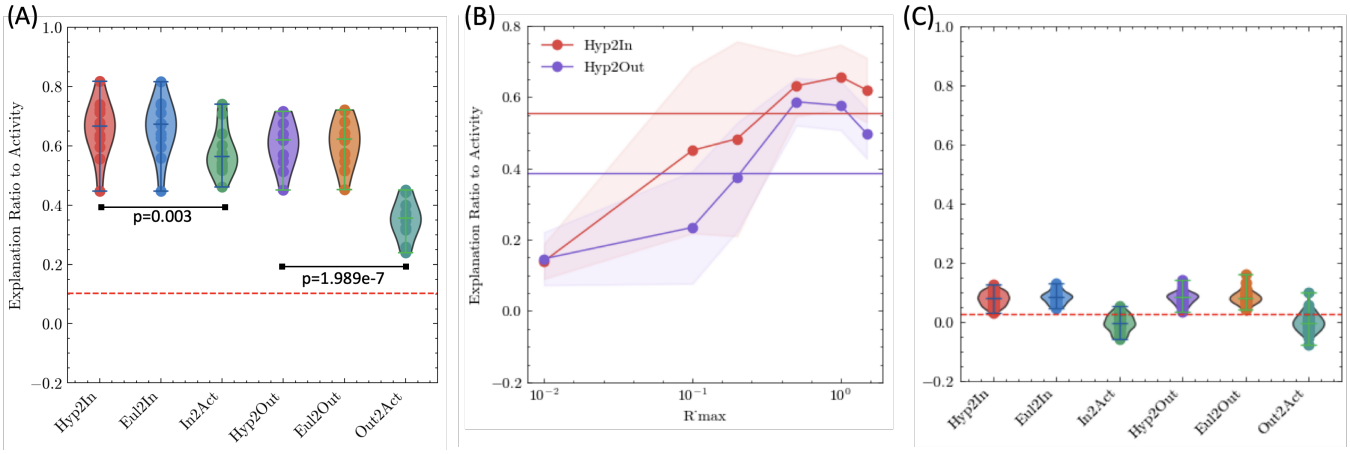

Fig. 3. Using Low-Dimensional Geometric Embeddings of Structural Relationship Improves Reconstruction Quality of Activity Traces: (A) Explanation ratio (normalized by scan-specific a_{act}) achieved by $\mathbf{C}^{\text{in}}$, $\mathbf{C}^{\text{out}}$, and their respective 2D Euclidean and hyperbolic embeddings ($R_{\max} = 1$). The dashed line (15.1%) indicates the average of the median pairwise neuronal activity correlations across 12 scans, normalized by a_{act}. The p-values comparing the performance of $\mathbf{C}^{\text{in}}$ and $\mathbf{C}^{\text{out}}$ with their corresponding hyperbolic embeddings are shown. Each scattered point corresponds to a separate scan and represents the median performance of scan-specific targeted neurons; (B) Sensitivity analysis of reconstruction accuracy as a function of embedding space capacity, varied through $R_{\max}$. Horizontal lines represent median accuracy for $\mathbf{C}^{\text{in}}$ and $\mathbf{C}^{\text{out}}$; (C): Similar analysis as in (A), but on an untrained continuous-time recurrent neural network, where no signal is represented through the connectome and minimal information is derived from its embeddings.

relationships in meso- and macro-scale brain networks (23). Additionally, using direct connectivity significantly limits the amount of data available for reconstruction and constrains performance, as the overall MICrONS dataset exhibits a sparsity

of only 0.12%, and the pairwise connections among our 497 targeted neurons have a sparsity of 1.75% (7). In contrast, the correlation measure, acting as a smooth filter, substantially broadens the scope of usable information and produces bounded and symmetric results, making it more suitable for embedding-based analyses. It also enables detailed characterization of preferential selectivity for both inputs and outputs separately (36, 39). Nevertheless, our approach does not address higher-order functional motifs or more complex organizational principles beyond pairwise rules, nor does it account for global topological features such as modularity, efficiency, or short path lengths (36, 75, 76).

Our findings demonstrate that weighted connectivity patterns yield activity prediction accuracy comparable to that achieved using binary connectivity, consistent with prior studies suggesting that the topological structure, defined by the adjacency matrix, is sufficient for structure-function prediction (36, 75). Also, the use of unweighted graphs to model excitable dynamics also offers computational advantages by simplifying network representation and analysis while maintaining high performance (77). Additionally, our results suggest that total synaptic strength, measured as the sum of PSD sizes, may not serve as a reliable indicator of such alignment (Figure 24). For improved interpretability, we employ a simple linear model to reconstruct activity traces. However, more complex neural-network-based models that incorporate nonlinearity and leverage embedding information, or that simply make $\boldsymbol{\alpha}$ parameters trainable, such as a single-layer perceptron, may yield improved results (31, 32, 64). Also, considering the design of neurophysiological experiments in which a head-fixed mouse freely walks on a cylindrical treadmill while viewing visual stimuli, the effectiveness of our simple model occurs may result from the passive nature of stimuli presentation, as the mouse is not engaged in any demanding task such as decision making based on contextual information or cues (Figure 31) (7, 20, 78). Evaluating the model's effectiveness and robustness under various experimental settings represents an interesting direction for future research. Additionally, we focus on activity reconstruction, noting that subtle modes of activity may play a critical role in behavior (79). Our work exclusively examines the recurrent circuit components reconstructed in the MICrONS dataset and does not make any assumptions or provide formulations regarding how stimuli are presented or transmitted through the sensory cortex (e.g. retinal ganglion cell (RGC) and lateral geniculate nucleus (LGN)) (36, 59, 80).

The effectiveness of using MDS to set weights in the linear model may be attributed to at least three factors. First, anatomical similarity may create a spatial embedding for each neuron, where similar neuron types are positioned closer together and exhibit similar activation patterns. This arises from the presence of multiple neuronal types, which are not explicitly differentiated in the recordings (81–83). Second, the spatial location of neurons imposes constraints on connection formation, which MDS then extracts from the connectome (8, 84–86). While the mouse cortex lacks organization in terms of orientation columns, it exhibits strong retinotopic organization (87, 88). These mechanisms are not independent, as spatial distance plays a key role in refining neuronal subtypes and predicting variations in synaptic connectivity patterns (39, 89). Third, neuronal responsiveness and activity patterns may be strongly influenced by the spatial and temporal organization of synaptic inputs, which, in conjunction with Hebbian plasticity, shape synchronous dynamics (40, 66, 90). MDS could effectively capture these dynamics. However, empirical measurements of the dimensionality of MICrONS activity space suggest that neural dynamics are not strictly low-dimensional (Figure 31), limiting the influence of this factor. A precise attribution of MDS effectiveness to these or other factors is beyond the scope of this study.

In conclusion, this study empirically quantifies the structure-function relationship using synaptic-scale connectivity correlations and demonstrates that low-dimensional embeddings of these correlations are more closely associated with neuronal dynamics. Specifically, detailed synaptic information from pairwise neuronal connections can be effectively represented as interactions in two- or three-dimensional spaces with varying geometry, enabled by improved reconstruction efficacy. Accordingly, we emphasize the concept of connectome embedding and its significance for understanding circuit-level mechanisms with broader implications. First, calculating input similarity using only excitatory neurons results in a 20.1% performance decrease, whereas the embedding approach incurs only a 3.11% reduction (Figure 3 and Figure 25). This robustness suggests that connectomic data remains highly informative even without a fully dense, though still unbiased, reconstruction. Second, the comparable performance of the embedded output highlights the potential presence of plasticity rules that generate network configurations amenable to low-dimensional embedding. Third, the effectiveness of connectome embedding indicates resolving every individual synapse may not be necessary to understand structure-activity relations. Alternatively, for example, one could construct a parsimonious, cell-type-based model capable of predicting neuronal connectivity, thereby significantly reducing computational costs in simulations.

ACKNOWLEDGEMENTS

We thank Kyle Aitken, Stefan Berteau, Eric Shea-Brown, Łukasz Kuśmierz, Zhixin Lu, Dana Mastrovito, Uygar Sümbül, and Praveen Venkatesh for discussions. We also wish to thank the Allen Institute for Brain Science founder, Paul G. Allen, for his vision, encouragement, and support.

1. Danielle S Bassett and Michael S Gazzaniga. Understanding complexity in the human brain. *Trends in cognitive sciences*, 15(5):200–209, 2011.
2. Michael Hawrylycz, Jeremy A Miller, Vilas Menon, David Feng, Tim Dolbeare, Angela L Guillozet-Bongaarts, Anil G Jegga, Bruce J Aronow, Chang-Kyu Lee, Amy Bernard, et al. Canonical genetic signatures of the adult human brain. *Nature neuroscience*, 18(12):1832–1844, 2015.
3. Frederico AC Azevedo, Ludmila RB Carvalho, Lea T Grinberg, José Marcelo Farfel, Renata EL Ferretti, Renata EP Leite, Wilson Jacob Filho, Roberto Lent, and Suzana Herculano-Houzel. Equal numbers of neuronal and nonneuronal cells make the human brain an isometrically scaled-up primate brain. *Journal of Comparative Neurology*, 513(5):532–541, 2009.
4. Nicole Hindley, Anna Sanchez Avila, and Christopher Henstridge. Bringing synapses into focus: Recent advances in synaptic imaging and mass-spectrometry for studying synaptopathy. *Frontiers in Synaptic Neuroscience*, 15:1130198, 2023.
5. Narayanan Kasthuri, Kenneth Jeffrey Hayworth, Daniel Raimund Berger, Richard Lee Schalek, José Angel Conchello, Seymour Knowles-Barley, Dongil Lee, Amelio Vázquez-Reina, Verena Kaynig, Thouis Raymond Jones, et al. Saturated reconstruction of a volume of neocortex. *Cell*, 162(3):648–661, 2015.
6. Thomas M Bartol Jr, Cailey Bromer, Justin Kinney, Michael A Chirillo, Jennifer N Bourne, Kristen M Harris, and Terrence J Sejnowski. Nanoconnectomic upper bound on the variability of synaptic plasticity. *Elife*, 4:e10778, 2015.
7. MICrONS Consortium, J Alexander Bae, Mahaly Baptiste, Caitlyn A Bishop, Agnes L Bodor,

Derrick Brittain, JoAnn Buchanan, Daniel J Bumbarger, Manuel A Castro, Brendan Celii, et al. Functional connectomics spanning multiple areas of mouse visual cortex. *Nature*, 640 (8058):435–447, 2025.
8. Ed Bullmore and Olaf Sporns. The economy of brain network organization. *Nature reviews neuroscience*, 13(5):336–349, 2012.
9. Christopher W Kuzawa, Harry T Chugani, Lawrence I Grossman, Leonard Lipovich, Otto Muzik, Patrick R Hof, Derek E Wildman, Chet C Sherwood, William R Leonard, and Nicholas Lange. Metabolic costs and evolutionary implications of human brain development. *Proceedings of the National Academy of Sciences*, 111(36):13010–13015, 2014.
10. Jared Kaplan, Sam McCandlish, Tom Henighan, Tom B Brown, Benjamin Chess, Rewon Child, Scott Gray, Alec Radford, Jeffrey Wu, and Dario Amodei. Scaling laws for neural language models. *arXiv preprint arXiv:2001.08361*, 2020.
11. Zahid Padamsey and Nathalie L Rochefort. Paying the brain's energy bill. *Current opinion in neurobiology*, 78:102668, 2023.
12. Yasaman Bahri, Ethan Dyer, Jared Kaplan, Jaehoon Lee, and Utkarsh Sharma. Explaining neural scaling laws. *Proceedings of the National Academy of Sciences*, 121(27): e2311878121, 2024.
13. Tanishq Kumar, Zachary Ankner, Benjamin F Spector, Blake Bordelon, Niklas Muennighoff, Mansheej Paul, Cengiz Pehlevan, Christopher Ré, and Aditi Raghunathan. Scaling laws for precision. *arXiv preprint arXiv:2411.04330*, 2024.
14. Mark M Churchland, John P Cunningham, Matthew T Kaufman, Justin D Foster, Paul Nuyujukian, Stephen I Ryu, and Krishna V Shenoy. Neural population dynamics during reaching. *Nature*, 487(7405):51–56, 2012.
15. John P Cunningham and Byron M Yu. Dimensionality reduction for large-scale neural recordings. *Nature neuroscience*, 17(11):1500–1509, 2014.
16. Ege Altan, Xuan Ma, Lee E Miller, Eric J Perreault, and Sara A Solla. Low-dimensional neural manifolds for the control of constrained and unconstrained movements. *bioRxiv*, pages 2023–05, 2023.
17. Emily R Oby, Alan D Degenhart, Erinn M Grigsby, Asma Motiwala, Nicole T McClain, Patrick J Marino, Byron M Yu, and Aaron P Batista. Dynamical constraints on neural population activity. *Nature Neuroscience*, pages 1–11, 2025.
18. Peter Dayan and Laurence F Abbott. *Theoretical neuroscience: computational and mathematical modeling of neural systems*. 2005.
19. Wulfram Gerstner, Marco Lehmann, Vasiliki Liakoni, Dane Corneil, and Johanni Brea. Eligibility traces and plasticity on behavioral time scales: experimental support of neohebbian three-factor learning rules. *Frontiers in neural circuits*, 12:53, 2018.
20. Christopher J Honey, Olaf Sporns, Leila Cammoun, Xavier Gigandet, Jean-Philippe Thiran, Reto Meuli, and Patric Hagmann. Predicting human resting-state functional connectivity from structural connectivity. *Proceedings of the National Academy of Sciences*, 106(6):2035–2040, 2009.
21. Andrea Avena-Koenigsberger, Bratislav Misic, and Olaf Sporns. Communication dynamics in complex brain networks. *Nature reviews neuroscience*, 19(1):17–33, 2018.
22. Bertha Vázquez-Rodríguez, Laura E Suárez, Ross D Markello, Golia Shafiei, Casey Paquola, Patric Hagmann, Martijn P Van Den Heuvel, Boris C Bernhardt, R Nathan Spreng, and Bratislav Misic. Gradients of structure–function tethering across neocortex. *Proceedings of the National Academy of Sciences*, 116(42):21219–21227, 2019.
23. Laura E Suárez, Ross D Markello, Richard F Betzel, and Bratislav Misic. Linking structure and function in macroscale brain networks. *Trends in cognitive sciences*, 24(4):302–315, 2020.
24. Zijin Gu, Keith Wakefield Jamison, Mert Rory Sabuncu, and Amy Kuceyeski. Heritability and interindividual variability of regional structure-function coupling. *Nature Communications*, 12 (1):4894, 2021.
25. Panagiotis Fotiadis, Linden Parkes, Kathryn A Davis, Theodore D Satterthwaite, Russell T Shinohara, and Dani S Bassett. Structure–function coupling in macroscale human brain networks. *Nature Reviews Neuroscience*, 25(10):688–704, 2024.
26. Fabian David Tschopp, Michael B Reiser, and Srinivas C Turaga. A connectome based hexagonal lattice convolutional network model of the drosophila visual system. *arXiv preprint arXiv:1806.04793*, 2018.
27. 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, 2020.
28. Lu Mi, Richard Xu, Sridhama Prakhya, Albert Lin, Nir Shavit, Aravinthan Samuel, and Srinivas C Turaga. Connectome-constrained latent variable model of whole-brain neural activity. In *International Conference on Learning Representations*, 2021.
29. Omer Mano, Matthew S Creamer, Bara A Badwan, and Damon A Clark. Predicting individual neuron responses with anatomically constrained task optimization. *Current Biology*, 31(18): 4062–4075, 2021.
30. Matthew S Creamer, Andrew M Leifer, and Jonathan W Pillow. Bridging the gap between the connectome and whole-brain activity in c. elegans. *bioRxiv*, pages 2024–09, 2024.
31. 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.
32. 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*, pages 1–9, 2024.
33. Gideon Rosenthal, František Váša, Alessandra Griffa, Patric Hagmann, Enrico Amico, Joaquín Goñi, Galia Avidan, and Olaf Sporns. Mapping higher-order relations between brain structure and function with embedded vector representations of connectomes. *Nature communications*, 9(1):2178, 2018.
34. Gidon Levakov, Joshua Faskowitz, Galia Avidan, and Olaf Sporns. Mapping individual differences across brain network structure to function and behavior with connectome embedding. *NeuroImage*, 242:118469, 2021.
35. Sven Dorkenwald, Casey M Schneider-Mizell, Derrick Brittain, Akhilesh Halageri, Chris Jordan, Nico Kemnitz, Manual A Castro, William Silversmith, Jeremy Maitin-Shephard, Jakob Troidl, et al. Cave: Connectome annotation versioning engine. *Nature Methods*, pages 1–9, 2025.
36. Zhuokun Ding, Paul G Fahey, Stelios Papadopoulos, Eric Y Wang, Brendan Celii, Christos Papadopoulos, Andersen Chang, Alexander B Kunin, Dat Tran, Jiakun Fu, et al. Functional connectomics reveals general wiring rule in mouse visual cortex. *Nature*, 640(8058):459–469, 2025.
37. Brendan Celii. *NEURD: automated proofreading and feature extraction for connectomics*. PhD thesis, Rice University, 2024.
38. Sen Song, Per Jesper Sjöström, Markus Reigl, Sacha Nelson, and Dmitri B Chklovskii. Highly nonrandom features of synaptic connectivity in local cortical circuits. *PLoS biology*, 3 (3):e68, 2005.
39. Rodrigo Perin, Thomas K Berger, and Henry Markram. A synaptic organizing principle for cortical neuronal groups. *Proceedings of the National Academy of Sciences*, 108(13): 5419–5424, 2011.
40. Ho Ko, Sonja B Hofer, Bruno Pichler, Katherine A Buchanan, P Jesper Sjöström, and Thomas D Mrsic-Flogel. Functional specificity of local synaptic connections in neocortical networks. *Nature*, 473(7345):87–91, 2011.
41. Lee Cossell, Maria Florencia Iacaruso, Dylan R Muir, Rachael Houlton, Elie N Sader, Ho Ko, Sonja B Hofer, and Thomas D Mrsic-Flogel. Functional organization of excitatory synaptic strength in primary visual cortex. *Nature*, 518(7539):399–403, 2015.
42. Andrew V Knyazev and Merico E Argentati. Principal angles between subspaces in an a-based scalar product: algorithms and perturbation estimates. *SIAM Journal on Scientific Computing*, 23(6):2008–2040, 2002.
43. Simon Kornblith, Mohammad Norouzi, Honglak Lee, and Geoffrey Hinton. Similarity of neural network representations revisited. In *International conference on machine learning*, pages 3519–3529. PMLR, 2019.
44. Ann E Sizemore, Jennifer E Phillips-Cremins, Robert Ghrist, and Danielle S Bassett. The importance of the whole: topological data analysis for the network neuroscientist. *Network Neuroscience*, 3(3):656–673, 2019.
45. Chad Giusti, Eva Pastalkova, Carina Curto, and Vladimir Itskov. Clique topology reveals intrinsic geometric structure in neural correlations. *Proceedings of the National Academy of Sciences*, 112(44):13455–13460, 2015.
46. Yuansheng Zhou, Brian H Smith, and Tatyana O Sharpee. Hyperbolic geometry of the olfactory space. *Science advances*, 4(8):eaaq1458, 2018.
47. Rishidev Chaudhuri, Berk Gerçek, Biraj Pandey, Adrien Peyrache, and Ila Fiete. The intrinsic attractor manifold and population dynamics of a canonical cognitive circuit across waking and sleep. *Nature neuroscience*, 22(9):1512–1520, 2019.
48. Richard J Gardner, Erik Hermansen, Marius Pachitariu, Yoram Burak, Nils A Baas, Benjamin A Dunn, May-Britt Moser, and Edvard I Moser. Toroidal topology of population activity in grid cells. *Nature*, 602(7895):123–128, 2022.
49. Huanqiu Zhang, P Dylan Rich, Albert K Lee, and Tatyana O Sharpee. Hippocampal spatial representations exhibit a hyperbolic geometry that expands with experience. *Nature Neuroscience*, 26(1):131–139, 2023.
50. Yuansheng Zhou and Tatyana O Sharpee. Hyperbolic geometry of gene expression. *Iscience*, 24(3), 2021.
51. Maximillian Nickel and Douwe Kiela. Poincaré embeddings for learning hierarchical representations. *Advances in neural information processing systems*, 30, 2017.
52. Frederic Sala, Chris De Sa, Albert Gu, and Christopher Ré. Representation tradeoffs for hyperbolic embeddings. In *International conference on machine learning*, pages 4460–4469. PMLR, 2018.
53. Marián Boguná, Fragkiskos Papadopoulos, and Dmitri Krioukov. Sustaining the internet with hyperbolic mapping. *Nature communications*, 1(1):62, 2010.
54. Tatyana O Sharpee. An argument for hyperbolic geometry in neural circuits. *Current opinion in neurobiology*, 58:101–104, 2019.
55. Jean-Baptiste Poline and Matthew Brett. The general linear model and fmri: does love last forever? *Neuroimage*, 62(2):871–880, 2012.
56. Trevor F Cox and Michael AA Cox. *Multidimensional scaling*. CRC press, 2000.
57. Carina Curto and Katherine Morrison. Relating network connectivity to dynamics: opportunities and challenges for theoretical neuroscience. *Current opinion in neurobiology*, 58:11–20, 2019.
58. Edgar Y Walker, Fabian H Sinz, Erick Cobos, Taliah Muhammad, Emmanouil Froudarakis, Paul G Fahey, Alexander S Ecker, Jacob Reimer, Xaq Pitkow, and Andreas S Tolias. Inception loops discover what excites neurons most using deep predictive models. *Nature neuroscience*, 22(12):2060–2065, 2019.
59. Eric Y Wang, Paul G Fahey, Zhuokun Ding, Stelios Papadopoulos, Kayla Ponder, Marissa A Weis, Andersen Chang, Taliah Muhammad, Saumil Patel, Zhiwei Ding, et al. Foundation model of neural activity predicts response to new stimulus types. *Nature*, 640(8058):470–477, 2025.
60. Patric Hagmann, Leila Cammoun, Xavier Gigandet, Reto Meuli, Christopher J Honey, Van J Wedeen, and Olaf Sporns. Mapping the structural core of human cerebral cortex. *PLoS biology*, 6(7):e159, 2008.
61. Carsten K Pfeffer, Mingshan Xue, Miao He, Z Josh Huang, and Massimo Scanziani. Inhibition of inhibition in visual cortex: the logic of connections between molecularly distinct interneurons. *Nature neuroscience*, 16(8):1068–1076, 2013.
62. Luke Campagnola, Stephanie C Seeman, Thomas Chartrand, Lisa Kim, Alex Hoggarth, Clare Gamlin, Shinya Ito, Jessica Trinh, Pasha Davoudian, Cristina Radaelli, et al. Local connectivity and synaptic dynamics in mouse and human neocortex. *Science*, 375(6585): eabj5861, 2022.
63. Xiaolong Jiang, Shan Shen, Cathryn R Cadwell, Philipp Berens, Fabian Sinz, Alexander S Ecker, Saumil Patel, and Andreas S Tolias. Principles of connectivity among morphologically defined cell types in adult neocortex. *Science*, 350(6264):aac9462, 2015.
64. Guangyu Robert Yang and Xiao-Jing Wang. Artificial neural networks for neuroscientists: a primer. *Neuron*, 107(6):1048–1070, 2020.
65. Peiran Gao, Eric Trautmann, Byron Yu, Gopal Santhanam, Stephen Ryu, Krishna Shenoy, and Surya Ganguli. A theory of multineuronal dimensionality, dynamics and measurement. *BioRxiv*, page 214262, 2017.
66. Ho Ko, Thomas D Mrsic-Flogel, and Sonja B Hofer. Emergence of feature-specific connectivity

in cortical microcircuits in the absence of visual experience. *Journal of Neuroscience*, 34 (29):9812–9816, 2014.
67. Adrian Wertz, Stuart Trenholm, Keisuke Yonehara, Daniel Hillier, Zoltan Raics, Marcus Leinweber, Gergely Szalay, Alexander Ghanem, Georg Keller, Balázs Rózsa, et al. Single-cell–initiated monosynaptic tracing reveals layer-specific cortical network modules. *Science*, 349(6243):70–74, 2015.
68. Wei-Chung Allen Lee, Vincent Bonin, Michael Reed, Brett J Graham, Greg Hood, Katie Glattfelder, and R Clay Reid. Anatomy and function of an excitatory network in the visual cortex. *Nature*, 532(7599):370–374, 2016.
69. M Florencia Iacaruso, Ioana T Gasler, and Sonja B Hofer. Synaptic organization of visual space in primary visual cortex. *Nature*, 547(7664):449–452, 2017.
70. L Federico Rossi, Kenneth D Harris, and Matteo Carandini. Spatial connectivity matches direction selectivity in visual cortex. *Nature*, 588(7839):648–652, 2020.
71. Benjamin Scholl, Connon I Thomas, Melissa A Ryan, Naomi Kamasawa, and David Fitzpatrick. Cortical response selectivity derives from strength in numbers of synapses. *Nature*, 590(7844):111–114, 2021.
72. Donald Olding Hebb. *The organization of behavior: A neuropsychological theory*. Psychology press, 2005.
73. Kelly Shen, Gleb Bezgin, R Matthew Hutchison, Joseph S Gati, Ravi S Menon, Stefan Everling, and Anthony R McIntosh. Information processing architecture of functionally defined clusters in the macaque cortex. *Journal of Neuroscience*, 32(48):17465–17476, 2012.
74. Maxwell H Turner, Kevin Mann, and Thomas R Clandinin. The connectome predicts resting-state functional connectivity across the drosophila brain. *Current biology*, 31(11):2386–2394, 2021.
75. Kerem Uzel, Saul Kato, and Manuel Zimmer. A set of hub neurons and non-local connectivity features support global brain dynamics in c. elegans. *Current Biology*, 32(16):3443–3459, 2022.
76. Mikail Rubinov and Olaf Sporns. Complex network measures of brain connectivity: uses and interpretations. *Neuroimage*, 52(3):1059–1069, 2010.
77. Arnaud Messé, Marc-Thorsten Hütt, and Claus Christian Hilgetag. Excitable dynamics simplify neural connectomes. *Cell Reports Physical Science*, 2025.
78. International Brain Laboratory, Valeria Aguillon-Rodriguez, Dora Angelaki, Hannah Bayer, Niccolo Bonacchi, Matteo Carandini, Fanny Cazettes, Gaelle Chapuis, Anne K Churchland, Yang Dan, et al. Standardized and reproducible measurement of decision-making in mice. *Elife*, 10:e63711, 2021.
79. Carsen Stringer, Marius Pachitariu, Nicholas Steinmetz, Charu Bai Reddy, Matteo Carandini, and Kenneth D Harris. Spontaneous behaviors drive multidimensional, brainwide activity. *Science*, 364(6437):eaav7893, 2019.
80. Daniel LK Yamins and James J DiCarlo. Using goal-driven deep learning models to understand sensory cortex. *Nature neuroscience*, 19(3):356–365, 2016.
81. Hongkui Zeng and Joshua R Sanes. Neuronal cell-type classification: challenges, opportunities and the path forward. *Nature Reviews Neuroscience*, 18(9):530–546, 2017.
82. Bosiljka Tasic, Zizhen Yao, Lucas T Graybuck, Kimberly A Smith, Thuc Nghi Nguyen, Darren Bertagnolli, Jeff Goldy, Emma Garren, Michael N Economo, Sarada Viswanathan, et al. Shared and distinct transcriptomic cell types across neocortical areas. *Nature*, 563(7729): 72–78, 2018.
83. Hongkui Zeng. What is a cell type and how to define it? *Cell*, 185(15):2739–2755, 2022.
84. Marc Barthélemy. Spatial networks. *Physics reports*, 499(1-3):1–101, 2011.
85. H Francis Song, Henry Kennedy, and Xiao-Jing Wang. Spatial embedding of structural similarity in the cerebral cortex. *Proceedings of the National Academy of Sciences*, 111(46): 16580–16585, 2014.
86. Jascha Achterberg, Danyal Akarca, DJ Strouse, John Duncan, and Duncan E Astle. Spatially embedded recurrent neural networks reveal widespread links between structural and functional neuroscience findings. *Nature Machine Intelligence*, 5(12):1369–1381, 2023.
87. Marina E Garrett, Ian Nauhaus, James H Marshel, and Edward M Callaway. Topography and areal organization of mouse visual cortex. *Journal of Neuroscience*, 34(37):12587–12600, 2014.
88. Jun Zhuang, Lydia Ng, Derric Williams, Matthew Valley, Yang Li, Marina Garrett, and Jack Waters. An extended retinotopic map of mouse cortex. *elife*, 6:e18372, 2017.
89. Lijuan Liu, Zhixi Yun, Linus Manubens-Gil, Hanbo Chen, Feng Xiong, Hongwei Dong, Hongkui Zeng, Michael Hawrylycz, Giorgio A Ascoli, and Hanchuan Peng. Neuronal connectivity as a determinant of cell types and subtypes. *Research Square*, pages rs–3, 2023.
90. Britton A Sauerbrei, Jian-Zhong Guo, Jeremy D Cohen, Matteo Mischiati, Wendy Guo, Mayank Kabra, Nakul Verma, Brett Mensh, Kristin Branson, and Adam W Hantman. Cortical pattern generation during dexterous movement is input-driven. *Nature*, 577(7790):386–391, 2020.
91. Leila Elabbady, Sharmishtaa Seshamani, Shang Mu, Gayathri Mahalingam, Casey Schneider-Mizell, Agnes L Bodor, J Alexander Bae, Derrick Brittain, JoAnn Buchanan, Daniel J Bumbarger, et al. Perisomatic features enable efficient and dataset wide cell-type classifications across large-scale electron microscopy volumes. 2022.
92. Robert Ghrist. Barcodes: the persistent topology of data. *Bulletin of the American Mathematical Society*, 45(1):61–75, 2008.
93. W Johnson J Lindenstrauss and J Johnson. Extensions of lipschitz maps into a hilbert space. *Contemp. Math*, 26(189-206):2, 1984.
94. William B Johnson and Assaf Naor. The johnson–lindenstrauss lemma almost characterizes hilbert space, but not quite. *Discrete & Computational Geometry*, 43(3):542–553, 2010.

A: Supplementary Materials

A. Details of MICrONS Dataset and Experiments. In this paper, all experiments and analysis are conducted using the MICrONS minnie65 V943 dataset, which contains 205,277,810 synapses and 90,230 neurons (7). The overall sparsity of pairwise connectivity is 0.127%. Among the recorded cells, 63,758 are excitatory neurons, 7,849 are inhibitory neurons, and 18,693 are non-neuronal cells, primarily glial cells. Regionally, 57,781 cells are located in V1, 21,184 in RL, and 10,480 in AL.

Proofreading plays a critical role in improving the quality of automated EM segmentation by resolving incorrect associations between presynaptic and postsynaptic partners. This process primarily addresses two types of reconstruction errors: false merges and false splits. False merges occur when segmented objects are erroneously grouped (e.g., an axon or dendrite incorrectly assigned to a soma), while false splits arise when objects are mistakenly separated (e.g., an axon or dendrite omitted from its corresponding soma). To correct false merges, the reconstruction is *cleaned* by removing incorrectly associated segments, whereas false splits are addressed by *extending* the reconstruction to restore missing segments. A detailed protocol for morphological proofreading is provided in (7, 36).

For axonal proofreading, 321 neurons are classified as extended, 714 as clean, and the remaining neurons are unproofread (91). Similarly, for dendritic proofreading, 44,163 neurons have been completely merged and properly cleaned – either manually or using the NEURD algorithm – resulting in fully reconstructed dendrites, while the remaining neurons are unproofread (37, 91). To increase the number of axonally trustworthy neurons, we relax the selection criterion to include neurons labeled as either extended or clean. In contrast, reconstructed dendrites are generally complete and require minimal extension, resulting in a greater number of neurons classified as dendritically trustworthy. However, only 845 neurons (fewer than 1% of the total) meet the structural confidence criteria for both axonal and dendritic aspects. In analyzing the geometric representation of the connectome (Figure G), we will consider different subsets, including all excitatory neurons ($N = 716$) and excitatory neurons in V1 ($N = 675$). For the analyses in Section 2 and the baseline experiment in Section 4, $\mathbf{C}^{\text{out}}$ is computed using connections exclusively *to* E/I neurons with fully reconstructed dendrites, whereas $\mathbf{C}^{\text{in}}$ is computed using connections exclusively *from* E/I neurons with extended or clean axons. Table 1 details the dimensionality of $\mathbf{W}^{\text{in}}$ and $\mathbf{W}^{\text{out}}$ under different experimental settings, and Tables 2 and 3 detail the cell identities. Neurons that are isolated in the connectivity are excluded, as their inclusion would result in entire rows and/or columns of NaN values in $\mathbf{C}$ (corresponding to 0 values in $\mathbf{W}$). Additional comparative studies, including those involving different connectome types $\mathbf{W}$ or the inclusion of synapses *to* or *from* unproofread neurons and glial cells, are discussed in Section 4 (Table 1, "Noise" row).

In this work, we focus on aligning the anatomical and functional properties of excitatory neurons that are both function-

ally coregistered (i.e. located within the volume of dense reconstruction, also see Figure 10) and structurally proofreaded. These selection criteria are implemented using CAVE (35). While restricting analysis to neurons that meet both criteria enhances data reliability, it significantly reduces the number of available neurons. Table 4 provides a detailed comparison of selection criterion (36). We use the number of neurons listed under "Number of **Coreg + Prf** Neurons" in the linear model presented in Section 4. Across all scans, only 497 **Coreg + Prf** neurons (excluding duplicate appearances) meet these criteria.

All the analysis code from this paper can be found on GitHub[1].

	Number of Input	Number of Output
Baseline	1035	25497
E Only	850	23019
I Only	185	2478
Noise	71607	71607

Table 1. Number of Input and Output Neurons Used in Calculating $\mathbf{C}^{\text{in}}$ and $\mathbf{C}^{\text{out}}$: Corresponds to the values of M_1, M_2. All rows, except for "Noise", include only neurons that are proofread on the axon side (for input) and the dendrite side (for output) correspondingly.

	L1	L2-3	L4	L5	L6	WM
Input	10	407	355	182	77	4
Output	402	7751	7453	3828	4719	1344
Targeted	1	205	157	92	42	0

Table 2. Layer Classification of Input, Output, and Target Neurons in the Baseline Experiment: L: Layer; WM: White Matter.

	V1	RL	AL	Others
Input	899	88	48	0
Output	18895	6158	413	31
Targeted	382	68	47	0

Table 3. Region Classification of Input, Output, and Target Neurons in the Baseline Experiment: V1: Primary Visual Cortex; AL: Anterolateral Area; RL: Rostrolateral Area.

B. More Scans of "Correlation of Correlation" Analysis.

See Figure 4.

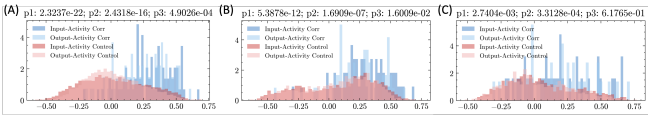

Fig. 4. Statistical Comparison of On-Diagonal and Off-Diagonal Elements of "Correlation of Correlation" Matrix: Expanding on Figure 1E1, we present results from additional scans to demonstrate the consistency and robustness of our observations. (A): Session 4, Scan 7; (B) Session 9, Scan 4; (C): Session 5, Scan 3. The p_1 to p_3 values, defined identically to those in Figure 1E1, are provided in the captions.

C. Subspace Angle Analysis. See Figure 5.

[1]https://github.com/StevenZhang0116/StructureFunctionAlignment

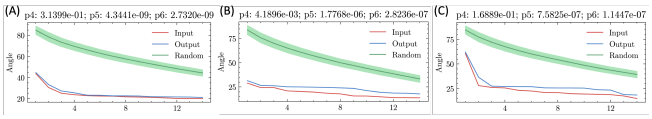

Fig. 5. Principal Angle Analysis: Expanding on Figure 1E2, we compute the principal angle between subspaces (flats) constructed by $\mathbf{C}^{\text{act}}$ and $\mathbf{C}^{\text{in}}$ (red), $\mathbf{C}^{\text{out}}$ (blue), and random matrices (green) (42). Statistical significance is indicated by p_4 for the difference between the red and blue curves, p_5 for the red and green curves, and p_6 for the blue and green curves. As discussed in the main text, the input affinity subspace exhibits the strongest alignment with the activity space. (A) Session 4, Scan 7; (B) Session 9, Scan 4; and (C) Session 8, Scan 5 (Figure 1).

D. Neuron Pairs with Direct Connectivity has Stronger Correlation in Activity. See Figure 6.

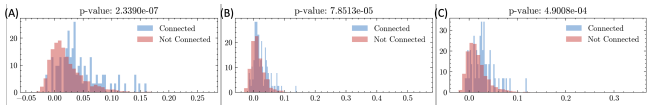

Fig. 6. Neurons with Strong Activity Correlation Are More Likely to Form Synaptic Connections: Unlike Figure 1, Figure 4, and Figure 5, which use non-local connectivity measures ($\mathbf{C}^{\text{in}}$ and $\mathbf{C}^{\text{out}}$) to probe the structure-function relationship, here we analyze the direct synaptic connectivity matrix between targeted neurons ($\mathbf{W} \in \mathbb{R}^{N \times N}$) and confirm the like-to-like principle (36). (A) Session 4, Scan 7; (B) Session 9, Scan 4; and (C) Session 8, Scan 5.

E. Clique Topology Method for Uncovering Hidden Geometries in Connectomic and Activity Structures.

It has been extensively discussed in previous literature that the clique topology method could be used to detect the geometric organization of the correlation or connectivity matrix, including purely random or embedded in Euclidean or hyperbolic space (45, 46, 49). In this analysis, we extend these discussions by using the input as a symmetric matrix $\mathbf{C}$, which quantifies the pairwise correlation between neurons i and j. The algorithm applies a step function to the matrix at various thresholds, retaining only the entries that exceed the specified value and generating corresponding topological graphs. These graphs are characterized by the number of cycles (holes) in different dimensions (1, 2, or higher). These counts, known as Betti numbers $\bar{\beta}_n$, are generally small at low thresholds (fully connected networks) or high thresholds (sparse networks where most nodes are not interconnected) (92). Plotting them as a function over the entire density range ρ of links (from the highest threshold to the lowest threshold) produces Betti curves $\beta_n(\rho), 0 \leq \rho \leq 1$. Clearly:

$$\bar{\beta}_m = \int_0^1 \beta_m(\rho) d\rho. \quad \textbf{(4)}$$

Typically, the top three Betti curves are sufficient to identify the underlying structures, and our analysis also follows this approach.

Uniformly sampling points from an n-dimensional Euclidean space is straightforward through high-dimensional uniform distribution sampling. Previous literature has also addressed sampling from n-dimensional hyperbolic space (46, 49). In the polar coordinate setting, both the curvature K and the maximal radius $R_{\max}$ can alter the geometry of the space. Specifically, changing K only rescales the radial coordinates of the nodes without affecting the topological properties of the

Session & Scan	Number of Recorded Neurons	Number of **Coreg** Neurons	Number of **Coreg** + **Prf** Neurons
Session 4, Scan 7	8395	1018	82
Session 8, Scan 5	10579	1123	68
Session 6, Scan 4	8826	1152	59
Session 9, Scan 4	8415	1027	56
Session 6, Scan 6	8551	1052	54
Session 5, Scan 6	9254	997	54
Session 5, Scan 7	8862	1000	41
Session 6, Scan 2	8756	1021	39
Session 7, Scan 3	9296	743	38
Session 7, Scan 5	9034	631	39
Session 9, Scan 3	8661	1170	37
Session 5, Scan 3	8790	426	35

Table 4. Available Neurons under Different Selection Criteria, Categorized by Session and Scan: Coreg: Functionally coregistered. **Prf**: Proofread.

network. For simplicity, we fix $\zeta := \sqrt{-K} = 1$ and treat $R_{\max}$ as a hyperparameter. Each point is sampled with uniformly distributed angles and a radius $r \in [R_{\min}, R_{\max}]$, following the distribution $\rho(r) \sim \sinh^{(d-1)r}$, which is approximately proportional to $e^{(d-1)r}$ when $r \gg 1$. In (46), $R_{\min}$ is set to $0.9R_{\max}$ to constrain the sampled points to have similar radii (i.e. similar to torus-like shape in $\mathbb{R}^3$). Following the convention in (49), we set $R_{\min} = 0$, leading to a less constrained distribution. Due to this difference in setting, the Betti curves presented in our work are qualitatively more similar to those in (49) and differ from those in (46) (e.g. decreasing prominence of peaks as m increases). Consequentially, unlike the L^2 distance in Euclidean space, the metric in hyperbolic geometry adheres the hyperbolic law of cosines:

$$\cosh(\zeta x) = \cosh(\zeta r)\cosh(\zeta r') - \sinh(\zeta r)\sinh(\zeta r')\cos\Delta\theta, \quad \textbf{(5)}$$

where x is the hyperbolic distance, r, r' the corresponding radii (of the two points) from the origin, and $\Delta\theta$ is the angle in between (49). To incorporate noise into the simulation, we consider the magnitude of Gaussian noise ϵ, which is scaled based on the geometric distance $\bar{\mathbf{D}}$ calculated through Equation (5):

$$\mathbf{D}_{ji} = \bar{\mathbf{D}}_{ji} \cdot (1 + \epsilon \cdot \mathcal{N}(0,1)). \quad \textbf{(6)}$$

We use the negative distance, $\mathbf{C}_{ji} = \mathbf{C}_{ij} = -\mathbf{D}_{ji}$, to quantify the correlation between neuron i and j. Any other order-reversing operation, including the reciprocal, could also accomplish this transformation, but we follow the previous literature and use the inverse for its stability in scaling (46, 49). In the main text, we primarily use Pearson's correlation to calculate $\mathbf{C}$. However, any operation that produces bounded and symmetric results can be applied, including Jensen-Shannon divergence, Hellinger distance, Jaccard distance, or Czekanowski-Dice distance for quantifying connectome similarity, and cosine similarity for quantifying activity similarity (85). In Section I, we demonstrate the robustness of the low-dimensional embedding geometry by using cosine similarity as an alternative method to compute $\mathbf{C}$ for activities.

In all simulations, we set N to match the size of each distance matrix as the input, with $\delta = 0.1$, $\epsilon = 0.0625$, $d = 2$ (or 3, see Section J), while fine-tuning the hypodermic radius $R_{\max}$ to achieve optimal alignment of the top three Betti curves. The optimization objective is to minimize the cumulative absolute pointwise error (Frobenius norm) of Betti curves using experimental and synthetic data. We validate the results by comparing the Betti number, $\bar{\beta}_m$. The analysis will display $\pm$ 1 standard deviation in all such cases. Through experiments, we observe that both the shape and magnitude of $\beta_m(\rho)$ are highly dependent on ϵ, which can be interpreted as representing various levels of mixing with random matrices (Equation 6). The parameter values are selected based on prior literature with minor adjustments (46, 49). Benefiting from recent developments[2], the scope of network size considered in the topological analysis presented in this paper surpasses that of previous works multiple times (46, 49).

A key difference between clique topology analysis and hyperbolic MDS[3] (Section 4) is that the latter applies dimensionality reduction to empirical data by minimizing the stress function and does not guarantee a uniform distribution of embedded points. The validity of our topological analysis is supported by the relatively uniform distribution of the targeted neurons, as confirmed through Euclidean non-metric MDS, which satisfies the assumptions underlying this analysis (Figure 7) (46, 49). In practice, we initialize the hyperbolic MDS algorithm with the converged Euclidean embedding of the same dimension and observe improvements in stress, highlighting the enhanced embedding accuracy of hyperbolic MDS.

F. Clique Topology Pipeline.

See Figure 8.

G. Representation Space of Neuronal Connectome is Low-Dimensional Hyperbolic Embedded.

See Figure 9, 10, 11, 12, and 13.

[2]https://github.com/nerdnik/PyCliqueTop_2023
[3]https://github.com/gyrheart/Hyperbolic-MDS

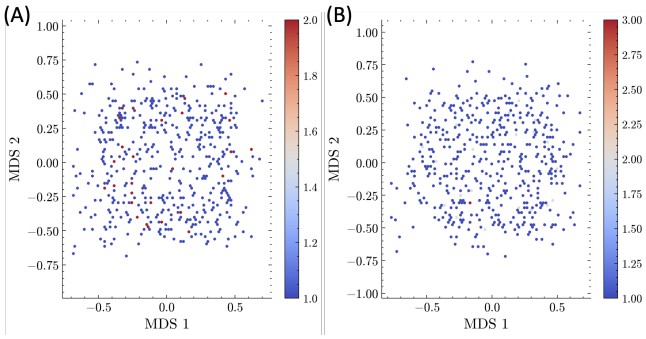

Fig. 7. 2D Euclidean MDS Embeddings of Input and Output Dissimilarity in the Baseline Experiment: The dissimilarity is calculated through 1 - Pearson correlation. Colored by frequency.

H. Additional Betti Curve Analysis on Representation Space of Activities Traces for All Scans. See Figure 16 and 17.

I. Betti Curve Analysis on Representation Space of Activities Traces Using Cosine Similarity. See Figure 18 and 19.

J. Equivalence Between 2D and 3D Hyperbolic Embeddings. See Figure 20. A similar discussion in Euclidean space is well addressed through the classical Johnson–Lindenstrauss lemma, which establishes the existence of a linear map $f : \mathbb{R}^N \to \mathbb{R}^n, N > n$, that nearly preserves the distances between points (93, 94). Here, we shift the focus from pairwise distance preservation to topological feature preservation and observe a perfectly linear relationship in the hyperbolic radius under pointwise transformation from $\mathbb{H}^3$ to $\mathbb{H}^2$. Our goal is not to prove the universality of such transformations but rather to highlight the embeddings in $\mathbb{H}^2$ discussed previously, including Section G, H, I, have direct counterparts in $\mathbb{H}^3$, as is typical in prior literature (46, 49). The range of $R_{\max}^{d=2}$ shown in Figure 20 aligns with the discussions in earlier sections.

K. Comparison of Betti Curves Between Random Matrices and Distance Matrices Sampled from Euclidean and Hyperbolic Geometries. See Figure 21.

L. Inclusion of More Neurons Enhances Activity Reconstruction Accuracy. See Figure 22.

M. Inclusion of More Neurons Does Not Improve the Explanation Ratio. See Figure 23.

N. Calculate Structural Similarity using Binary Connectome or Postsynaptic Density (PSD) Information. See Figure 24.

O. Calculate Structural Similarity Exclusively with Excitatory and Inhibitory Neurons. See Figure 25.

P. Calculate Structural Similarity using Non-fully-proofread Neurons. See Figure 26.

Q. Calculate Functional Similarity per Visual Stimulus Period. See Figure 27.

R. Reconstruction of Raw Fluorescence Traces. See Figure 28 and Table 5.

	Raw Trace	Deconvolved Trace
HypEmbed $\mathbf{C}^{\text{in}}$	24.1%	8.66%
$\mathbf{C}^{\text{in}}$	23.1%	7.70%
HypEmbed $\mathbf{C}^{\text{out}}$	22.8%	8.05%
$\mathbf{C}^{\text{out}}$	15.8%	4.92%
$\mathbf{C}^{\text{act}}$	36.4%	13.4%
Mean of $\mathbf{C}^{\text{act}}$	7.00%	1.36%

Table 5. Performance (a) Comparison between Reconstructing Raw Fluorescence Traces and Detrended, Deconvolved Traces (Used in Main Text) via the Linear Model: Different forms of $\mathbf{C}$ are used to compute the coefficients $\boldsymbol{\alpha}$ under the baseline experiment setup (Section 4 and Figure 3).

S. Reconstruction using Partially or Randomly Selected Activity Traces. See Figure 29.

T. Reconstruction Only Using Activity from Proofread Neurons. See Figure 30.

U. Dimensionality of the Activity Space for Coregistered Neurons and Randomly Selected Recorded Neurons. See Figure 31.

V. Effect of Euclidean Embedding Dimensionality on Reconstruction. See Figure 32.

W. Using Direct Synaptic Connectivity in Linear Model. See Figure 33.

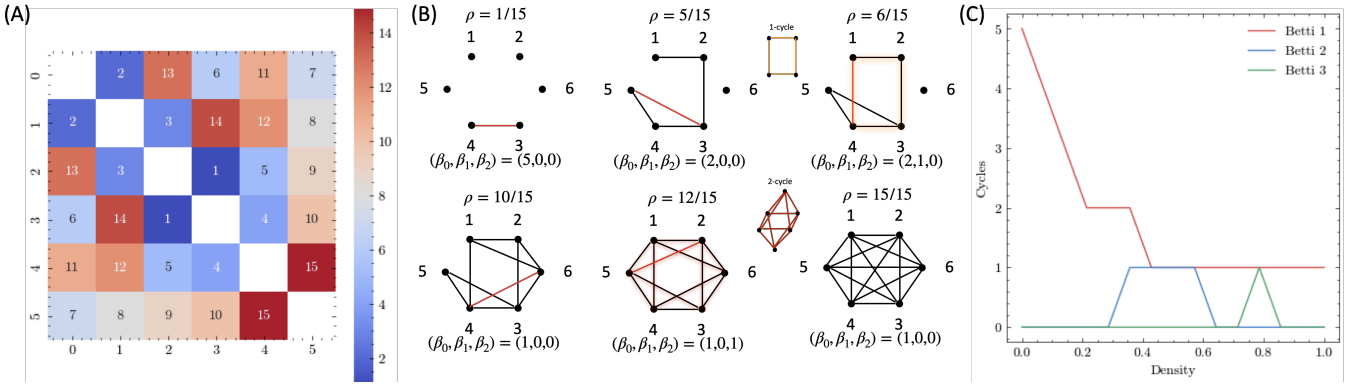

Fig. 8. Illustrative example of order-based analysis of symmetric matrics: (A): An arbitrary correlation matrix where each cell's value is converted to its relative rank; (B): By iteratively adjusting the threshold (filtration density), a sequence of binary adjacency matrics (and graphs) is generated. β_0 counts the number of connected components, while other β_m values count the number of m-cycles, such as 1-cycles (squares), 2-cycles (octahedra), and 3-cycles (orthoplex, not shown), which may be created, persist, and disappear at different densities. The colored shadow suggests the presence of a cycle; (C): The overall Betti curves $\beta_m(\rho)$ are shown by combing all the discrete instances discussed in (B). This example is reproduced from Dr. Nicole Sanderson's tutorial at IPAM's Mathematical Approaches for Connectome Analysis Workshop.

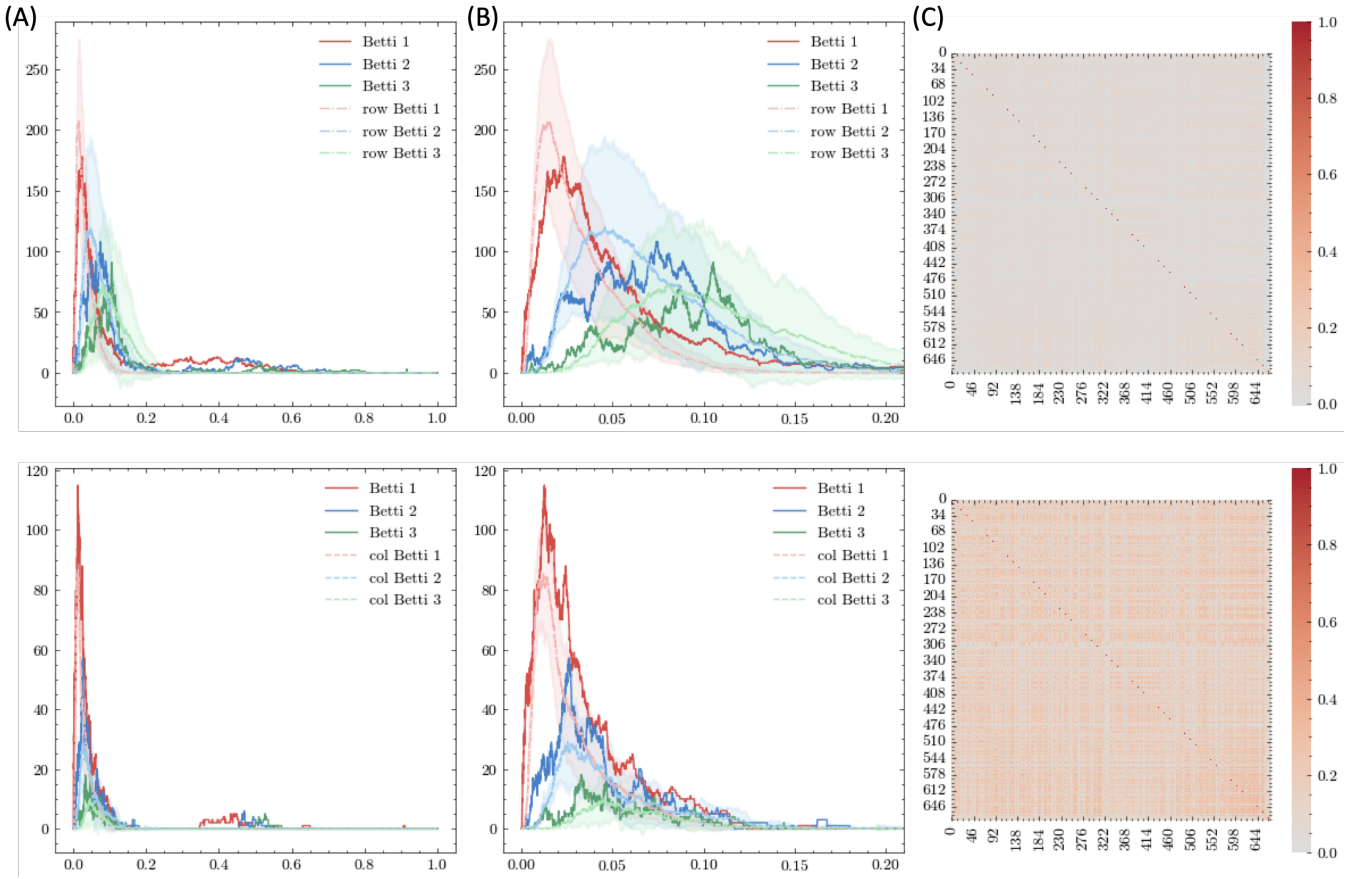

Fig. 9. 2D Hyperbolic Embedding of Structural Representation of Binary Connectome Among Proofread Excitatory Neurons in V1: To compare with Figure 1F in the main text, square Pearson correlation matrices of the binary connectome are computed for all proofread excitatory neurons in V1 ($N = 675$). As discussed in Section 4, structurally incomplete (unchecked) yet informative data is included. $\mathbf{C}^{\text{out}}$ (top row) and $\mathbf{C}^{\text{in}}$ (bottom row) are calculated using the rectangle matrix that includes all recorded excitatory and inhibitory neurons from MICrONS Version 943(7). The reconstructed neurons are distributed across two distinct cylindrical regions, corresponding to the V1 area (94.0%) and the RL/AL area (6.0%). Overall, their structural similarities, both for inputs and outputs, form two clusters, violating the assumption of uniform distribution in Betti analysis (49). Here, we focus specifically on V1 neurons; however, the results remain consistent for RL/AL neurons (Figure 13). A finer classification that groups neurons by cell types produces smoother and more regular Betti curves, thereby achieving better alignment with hyperbolic geometry in terms of topological feature mapping (Figure 14 and 15): (A): Betti curves for groundtruth and synthetic data, sampled from a 2D hyperbolic space (see Section E), averaged across 10 trials due to the high computational memory costs; results are shown with best fits via (pointwise) mean squared error (MSE) comparison to the mean curve (dashed lines) with optimized $R_{\max}$ of 18 and 14 respectively (grid size $= 1$); (B): Zoomed-in view of (A) highlighting the low edge density region; (C): Heatmap visualization of each correlation matrix. Using alternative correlation measures instead of Pearson correlation yields consistent results (not shown), as discussed in Section E and (85).

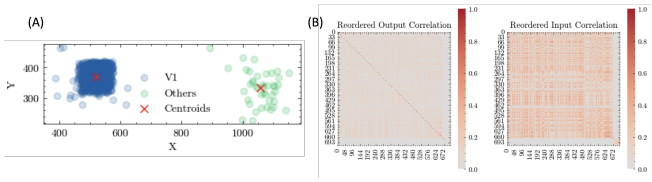

Fig. 10. Spatial Distribution of Reconstructed Neurons: (A): Cross-sectional projection on the XY plane (in μm) of all reconstructed neurons ($N = 716$), clustered using the K-means algorithm (7); (B): Sorted $\mathbf{C}^{\text{out}}$ and $\mathbf{C}^{\text{in}}$, arranged according to the labels in (A), and the matrices correspond to Figure 9.

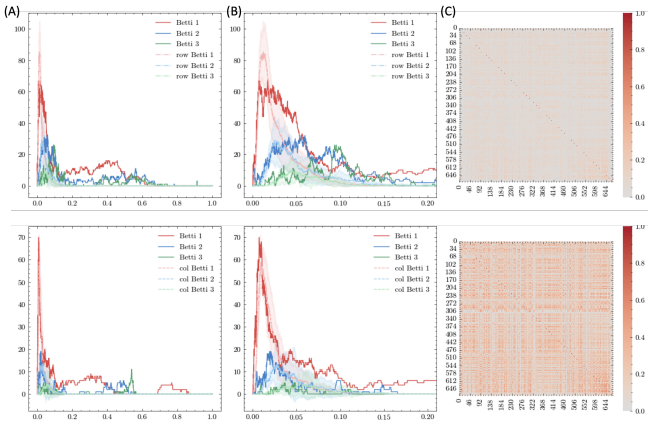

Fig. 11. 2D Hyperbolic Embedding of the Structural Representation of the Weighted Connectome Among Proofread Excitatory Neurons in V1: Similar to Figure 9, but $\mathbf{C}^{\text{out}}$ and $\mathbf{C}^{\text{in}}$ are computed using synaptic counts between neuronal connections, with optimized $R_{\max}$ of 14 and 13, respectively. As in Figure 9, the distinct emergence of a second peak in the high edge density region is consistent with observations in (46) and may be attributed to the significant nonuniform distribution of points in $\mathbb{H}^n$.

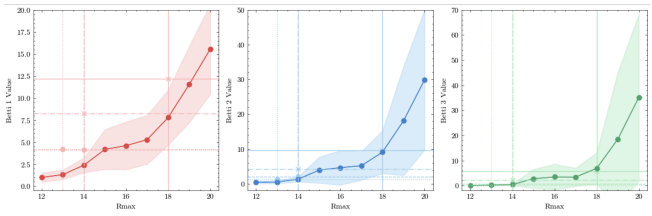

Fig. 12. Direct Comparison of Betti Values $\bar{\beta}_m$ Based On Groundtruth and Synthetic Data: Results derived from Figure 9 and Figure 11; Vertical lines indicate the optimized $R_{\max}$ values for $\mathbf{C}^{\text{out}}$ and $\mathbf{C}^{\text{in}}$ (4 in total, for binary and weighted connectome respectively); Horizontal lines represent the $\bar{\beta}_m$ values, $m = 1, 2, 3$, calculated based on the groundtruth. The deviations from the intersection points ("$*$" for $\mathbf{C}^{\text{out}}$ and "o" for $\mathbf{C}^{\text{in}}$) to the shadows indicate the effectiveness of approximations. All results are well-aligned except Betti 1 due to the occurrence of the second peaks at the high edge density region. The results further demonstrate the accuracy of the fits.

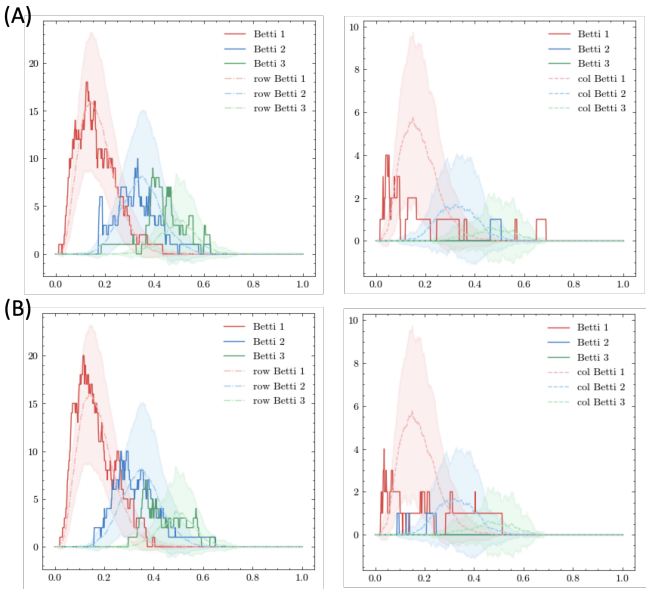

Fig. 13. 2D Hyperbolic Embedding of the Structural Representation of the Weighted Connectome Among Proofread Excitatory Neurons in RL/AL: Similar to Figure 9 and 11, this analysis focuses on neurons in higher visual areas (RL/AL) ($N = 41$) using (A) binary connectome or (B) weighted connectome to evaluate $\mathbf{C}^{\text{out}}$ (left column) and $\mathbf{C}^{\text{in}}$ (right column). Synthetic data are presented with optimized $R_{\max}$, averaged across 100 trials. Similar to observations in V1, the topological features of the representation space formed by $\mathbf{C}^{\text{in}}$ are less pronounced and can be approximated using a small $R_{\max}$, making it more similar to Euclidean space. This may be attributed to neuronal axons traveling much greater distances than dendrites, or alternatively, to an imbalance in reconstruction.

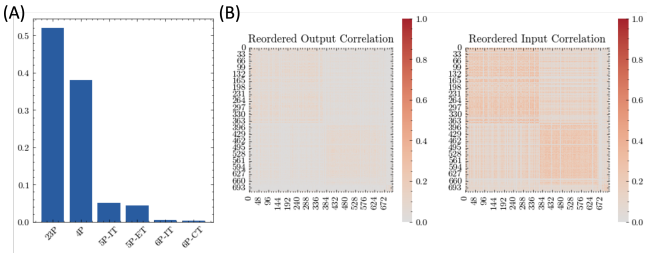

Fig. 14. Cell Type of Reconstructed Neurons: (A): Labeled cell types of all reconstructed neurons, where layer 2-3 pyramidal cells (23P, 52.0%) and layer 4 pyramidal cells (4P, 38.0%) are dominant (7); (B): Sorted $\mathbf{C}^{\text{out}}$ and $\mathbf{C}^{\text{in}}$, arranged according to the labels in (A).

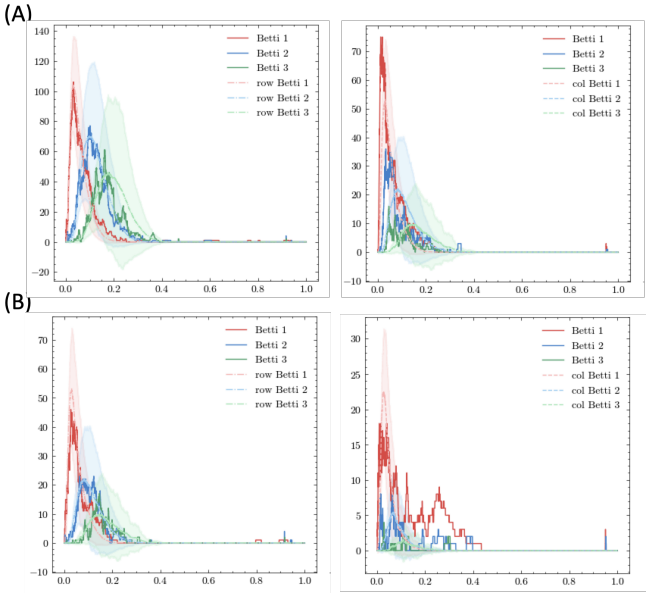

Fig. 15. 2D Hyperbolic Embedding of the Structural Representation Among Proofread Excitatory Neurons in 4P: Similar to Figure 13, this analysis focuses on layer 4 pyramidal cells ($N = 272$) using (A) binary connectome or (B) weighted connectome to evaluate $\mathbf{C}^{\text{out}}$ (left column) and $\mathbf{C}^{\text{in}}$ (right column). Synthetic data are presented with optimized $R_{\max}$, averaged across 100 trials.

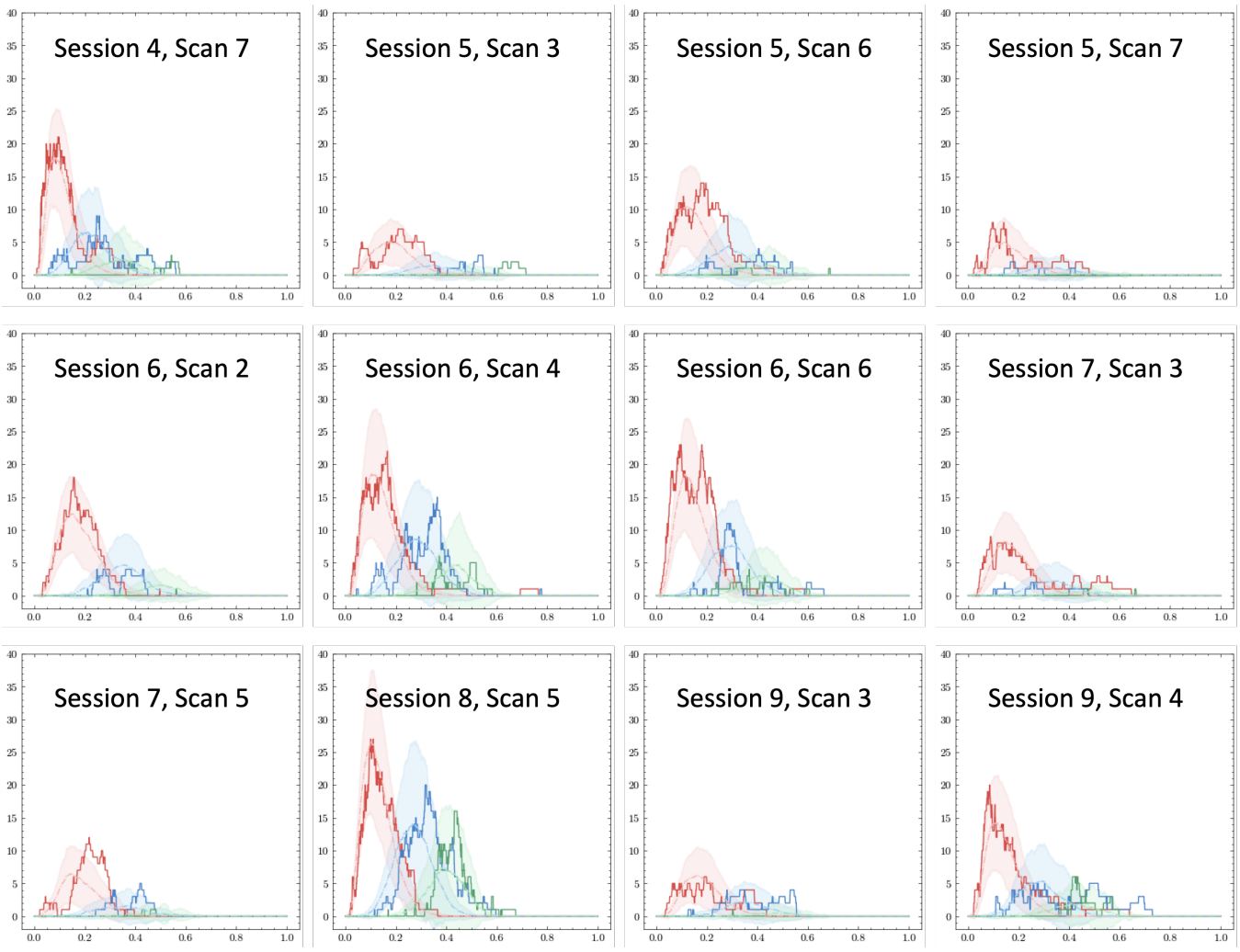

Fig. 16. Betti Analysis of All Activity Traces Indicates a Universal Hyperbolic Embedding: Repeat the analysis in Figure 1F for all scans by comparing the alignment of the top three Betti curves between groundtruth and synthetic data, averaged across 100 trials. Use the parameters described in Section E, with varying $R_{\max}$ values optimized within the range of 5 to 20 (grid size $= 0.1$).

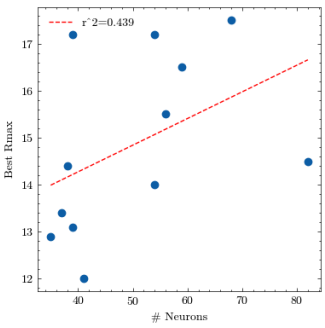

Fig. 17. Positive Relationship Between the Number of Neurons in the Scan and the Embedded Hyperbolic Radius: Results derived from Figure 16; including more neurons increase the effective dimensionality of the representation space, thereby necessitating greater capacity.

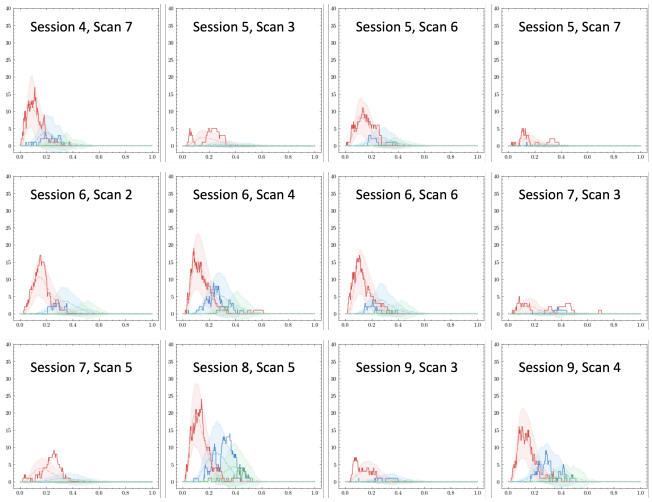

Fig. 18. Betti Analysis of All Activity Traces Using Cosine Similarity: Similar to Figure 16, but using cosine similarity to quantify activity proximity.

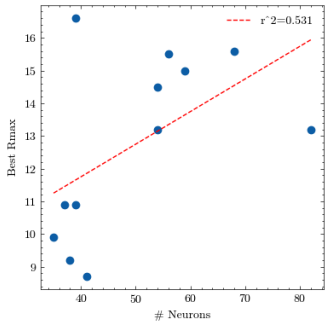

Fig. 19. Positive Relationship Between the Number of Neurons in the Scan and the Embedded Hyperbolic Radius Using Cosine Similarity as the Correlation Measure: Results derived from Figure 18; Similar to Figure 17. Under this measure, the representation space becomes less complex, reducing the magnitude of the Betti curves and an overall decrease in the hyperbolic radii $R_{\max}$.

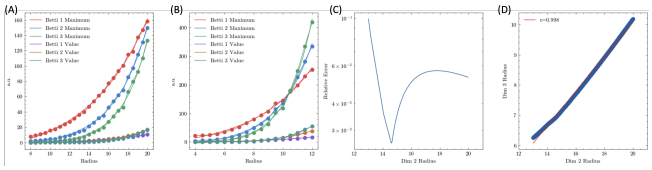

Fig. 20. Transformation between 2D and 3D hyperbolic embedding can be achieved through an approximately linear relationship of $R_{\max}$: Specific case for $N = 217$ and noise level $\epsilon = 0.0625$ averaged over 100 trials. (A-B): Presentations of $\bar{\beta}_m$ and $\max(\beta_m(\rho))$, $m = 1, 2, 3$ for 2D (A) and 3D (B). Each curve is interpolated using a simple exponential function and transformed into the continuous case; (C-D): Optimized cumulative pointwise error (C) and the mapping relationship (D) between $R^{d=2}_{\max}$ and $R^{d=3}_{\max}$. Note that when $R^{d=2}_{\max} \leq 13$, the $\bar{\beta}$ values are small (B part), which could lead to stability issues.

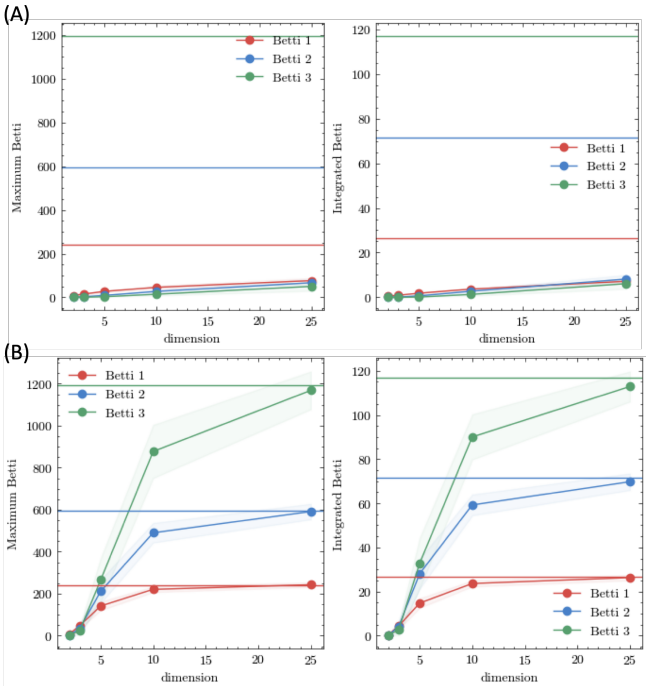

Fig. 21. Comparison of Random Matrices and Distances Matrices Sampled from Hyperbolic and Euclidean Spaces Across Various Dimensions: Variations in Betti values $\bar{\beta}_m$ (left column) and maximum Betti curve value $\max(\beta_m(\rho))$ (right column) for $m = 1, 2, 3$ are presented across different dimensions with $N = 100$ for (A) Euclidean space and (B) hyperbolic space, using the parameters specified in Section E. Horizontal lines indicate benchmark outcomes derived from random matrices. Each experiment is repeated for 500 trials.

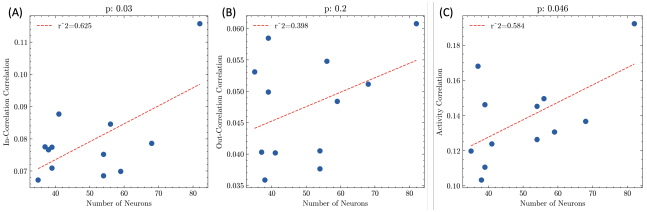

Fig. 22. Having More Functionally Coregistered Neurons in a Single Scan Enhances Reconstruction Accuracy: This result remains consistent when using (A) $\mathbf{C}^{\text{in}}$, (B): $\mathbf{C}^{\text{out}}$, and (C): $\mathbf{C}^{\text{act}}$. This example considers only proofread neurons with the weighted connectome, but the result is applicable to various settings.

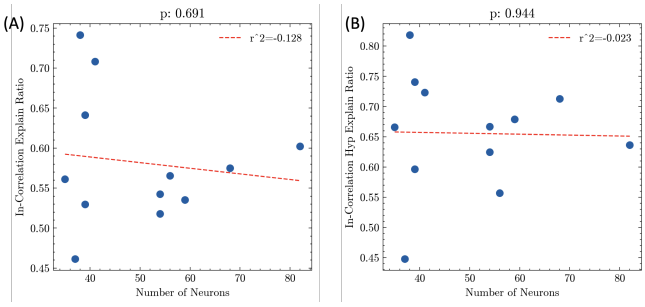

Fig. 23. Having More Functionally Coregistered Neurons in a Single Scan Does Not Improve the Explanation Ratio, Regardless of Full Connectome Information or Embedded Representation: Compared to Figure 22, although the absolute reconstruction accuracy a significantly improves with increasing N, the explanation ratio (a/a_{act}) remains uncorrelated with N for both the full connectome $\mathbf{C}^{\text{in}}$ (left) and its hyperbolic embeddings (right). The observation holds consistently for output correlation (not shown). Data points correspond to the green and red parts in Figure 3.

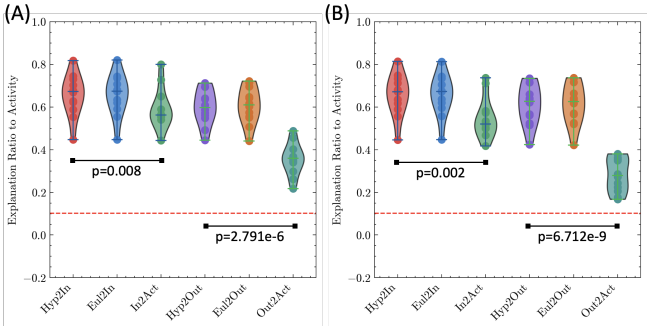

Fig. 24. Calculate $\mathbf{C}^{\text{in}}$ and $\mathbf{C}^{\text{out}}$ (and embeddings) using (A) binary connectome (A) or (B) postsynaptic density (PSD) information.

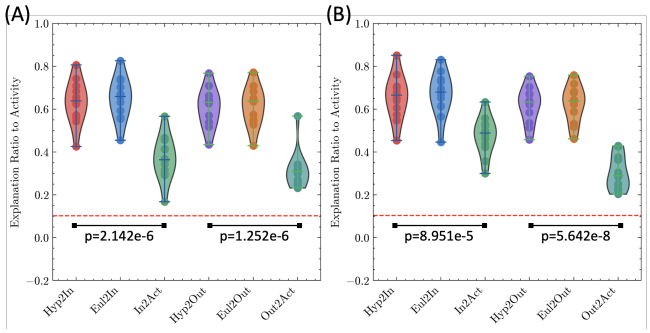

Fig. 25. Calculate $\mathbf{C}^{\text{in}}$ and $\mathbf{C}^{\text{out}}$ (and embeddings) using weighted connectome to/from only (A) excitatory neurons and (B) inhibitory neurons.

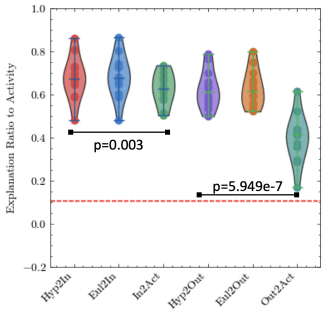

Fig. 26. Calculate $\mathbf{C}^{\text{in}}$ and $\mathbf{C}^{\text{out}}$ (and embeddings) using structural information from all excitatory and inhibitory neurons, including not fully proofread ones.

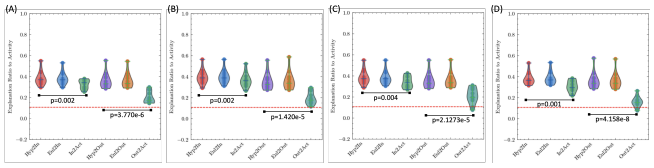

Fig. 27. Calculate $\mathbf{C}^{\text{act}}$ respectively for each visual stimulus period based on (A) synaptic count with proofread neurons, (B) synaptic count with all E/I neurons, (C) binary connectome with proofread neurons, and (D) PSD with proofread neurons to calculate structural similarity.

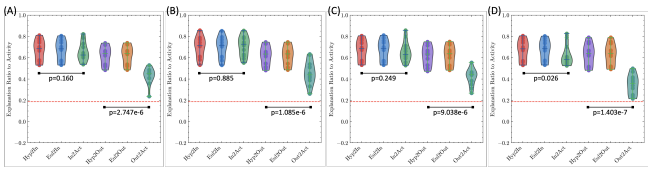

Fig. 28. Reconstruction of Raw Fluorescence Traces based on (A) synaptic count with proofread neurons, (B) synaptic count with all E/I neurons, (C) binary connectome with proofread neurons, and (D) PSD sizes with proofread neurons to calculate structural similarity: The explanation ratio based on the median pairwise neuronal activity increases to 18.6% (Figure 3).

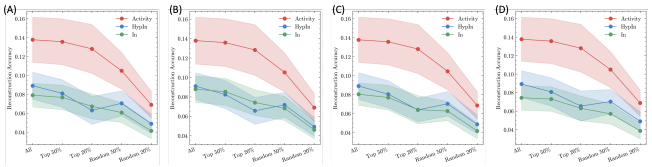

Fig. 29. Reconstruction using Partially or Randomly Selected Neuronal Activity Traces (averaged over 100 trials), based on (A) synaptic count with proofread neurons, (B) synaptic count with all E/I neurons, (C) binary connectome with proofread neurons, and (D) PSD sizes with proofread neurons to compute structural similarity. Means and standard deviations are calculated across 12 independent scans.

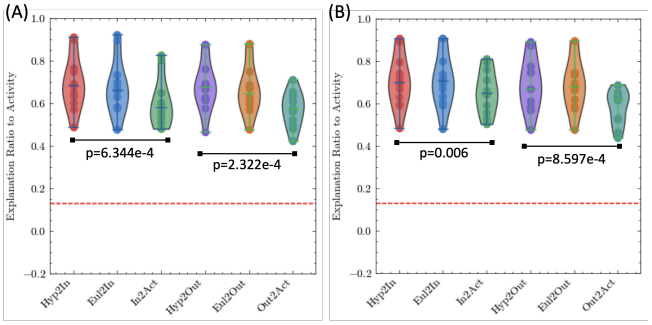

Fig. 30. Reconstruction using activity from proofread neurons based on (A) synaptic count with proofread neurons and (B) synaptic count with all E/I neurons to calculate structural similarity: The graph illustrates explanation ratio (a/a_{act}). As discussed in Section 4 and Figure 22, overall accuracies decrease (except the output reliability) due to the stricter criterion and less neurons included per scan. For (A), $\mathbf{C}^{\text{in}}$ achieves 6.78% and $\mathbf{C}^{\text{out}}$ achieves 6.50%, while the corresponding counterpart values are 7.17% and 4.65%, respectively. (B) has 7.28% and 6.65%, and the counterpart has 8.08% and 5.01%.

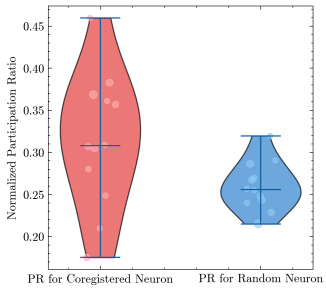

Fig. 31. Dimensionality of Activity Space: In the main text, we exclusively considered neurons that are functionally coregistered (i.e. neurons for which connectome information, though potentially incomplete, is available). We quantify the dimensionality of the activity space formed by these N_i neurons using the normalized participation ratio (NPR) (65). Additionally, we evaluate the NPR for randomly selected sets of N_i neurons in each scan, averaged over 100 instances. The size of the scattered points is scaled with N_i.

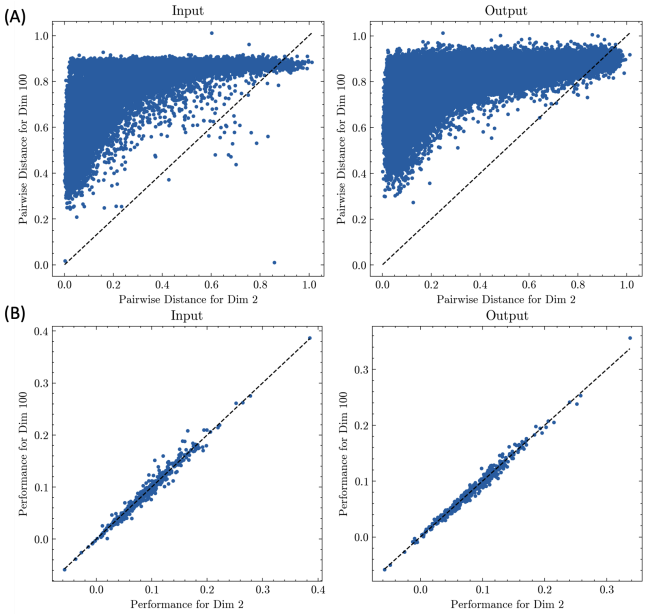

Fig. 32. Comparison of (A) pairwise distances between Euclidean embedded points and (B) neuron-wise reconstruction accuracy, concatenated across all scans, for different dimensionality: We compare the results using optimized Euclidean embeddings in $\mathbb{R}^2$ and $\mathbb{R}^{100}$. Despite significant variation in distance preservation quality, the neuron-wise reconstruction accuracy remains consistent. In (B), we averaged the performance (Pearson correlation) for neurons included in multiple scans. Compared to Section A, $\mathbb{R}^{100}$ would still be considered "low-dimensional" given the dataset size. Although it achieves near zero stress (not shown), the rank preservation (of pairwise distance remains imperfect ($R^2 = 0.73$ for the input and $R^2 = 0.92$ for the output), highlighting its distinction from the raw connectome data. Due to the computational challenges and instability caused by the high degree of freedom in "high"-dimensional embeddings, we limit the test to $\mathbb{R}^{100}$.

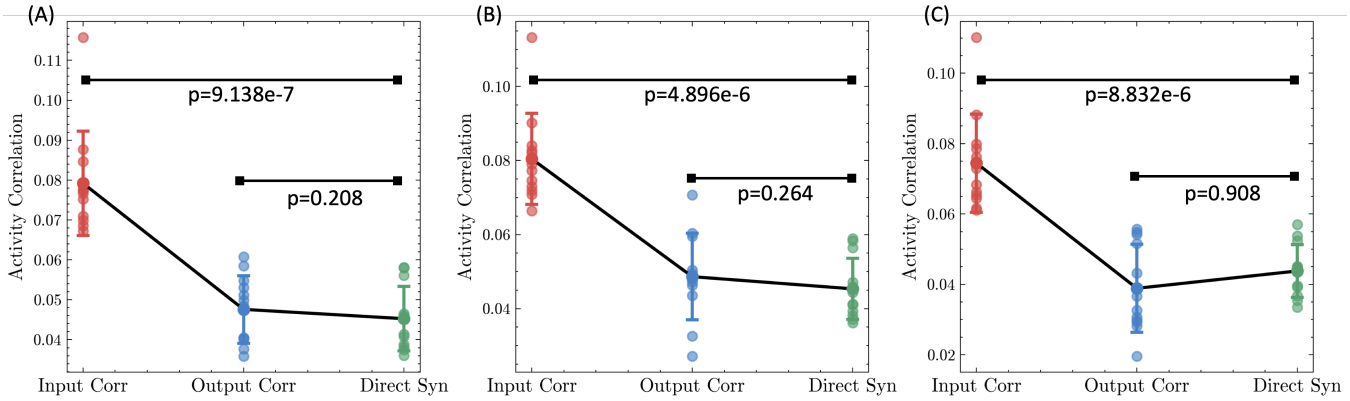

Fig. 33. Comparison Between Using Normalized Direct Connectivity as Coefficients in the Linear Model and Using Input and Output Similarity: We compare the reconstruction performance in Figure 3 and Figure 24, using input and output correlation respectively versus direct connectivity measures. The direct connectivity is represented through (A) synaptic count, (B) binary connectivity, and (C) PSD sizes. Means and standard deviations are reported.