

¹ Hebb's Vision: The Structural Underpinnings of
² Hebbian Assemblies

³ Full Author List TDB

⁴ **1 Abstract**

⁵ In 1949, Donald Hebb's theory of cell assemblies postulated that repeated functional
⁶ patterns could induce structural changes which group cells into organizational com-
⁷ ponents of perception and behavior. Hebb made influential predictions relating these
⁸ units' structure and activity, but it has remained challenging to validate his predictions
⁹ simultaneously. In this work, we leverage a novel, large-scale, multimodal dataset from
¹⁰ the murine visual cortex to test this relation. We extract cell assemblies from func-
¹¹ tional recordings by applying an existing algorithm that identifies assemblies as network
¹² communities based on moments of high coactivity. We find that assembly activity is
¹³ more reliable for repeated natural movie inputs than average cellular responses. Fur-
¹⁴ thermore, we show that these assemblies outperform other common functional groupings
¹⁵ in decoding responses to visual stimuli. In characterizing the structural underpinnings
¹⁶ of assemblies, we surprisingly observed no marked contrast in first-order connectivity
¹⁷ between cells with shared vs. disjoint assembly membership. However, higher-order
¹⁸ connection patterns revealed significant differences, such as stronger feed-forward in-
¹⁹ hibitory connections in disjoint memberships than in shared memberships. Additionally,
²⁰ we determined neurons outside these assemblies were significantly less integrated into
²¹ the structural network. Altogether, our results indicate that an assembly-based orga-
²² nization is an effective description of stimulus representation and results from mutual
²³ inhibitory organization.

24 2 Introduction

25 Since Hebb's 1949 monograph, "The Organization of Behavior" [25], Cell assemblies
26 have held a persistent place in the imagination of the neuroscience community, both as a
27 prospective unit of functional organization, and as a compellingly likely consequence of
28 simple synaptic plasticity rules, such as the 'fire-together wire-together' synapses which
29 were dubbed 'Hebbian Synapses' by Yves Frégnac in 1986. [15] The theory continues to
30 stimulate research into the activity-related aspects of assemblies, and these processing
31 modules' organization are thought to operate through stable recurrent activity ([10, 41]).
32 Incorporating a behavioral approach, experimental evidence implies that the response of
33 discrete cell populations akin to assemblies may have a causal link with motor functions
34 as well ([40]). In particular, there has been confirmation of sequential activation within
35 V1 neural assemblies ([11]) as well as ongoing 'replay' of coactivity in the absence of
36 stimuli ([36]). There also has been some limited evidence of cell assemblies stability over
37 long periods of time ([48]).

38 Research examining the structural aspects of Hebbian cell assemblies, however, has
39 primarily focused on potentiation of synapses between excitatory neurons. This is per-
40 haps not surprising, given that such reinforcement holds a place of prominence in Hebb's
41 original formulation of the theory [25]. Writing in 1963 [24], Hebb himself acknowledged
42 that his excitation-only formulation of assembly theory "had to be specific to an un-
43 palatable degree in order to show that a theory of neural connections in detail could
44 still, in 1949, be consistent with the known facts both of behavior and neurophysiol-
45 ogy." Critically, when it comes to inhibition, Hebb (*ibid*) states that "Brock, Coombs,
46 and Eccles (1952) [9] then presented decisive evidence, at last, of the cellular inhibition
47 which earlier would have meant one more arbitrary assumption and which consequently
48 I had to get along without." In other words, despite the entrenchment over the years
49 of an association between Hebbian assembly formation and Hebbian-style LTP, such
50 changes (and accompanying structural changes in synaptic knobs) were not a necessary
51 precondition for assembly formation. This line of thought has been continued [26], with
52 Holtmaat et al. (2016) positing that assemblies could theoretically form and function
53 primarily on the basis of a reinforcement of inhibition.

54 The approach of this work, using large-scale recordings of individual neurons to
55 identify and analyze assembly function, enters a developing tradition in the literature of
56 population dynamics [21] [22] [23] [30] [60] [65]. However, due to limitations in structural
57 analysis, few studies can relate the assemblies structure to function, as reviewed by [26].
58 Electrophysiological datasets can allow highly reliable inference of connectivity in the
59 case of multi-patch recordings, but with very small numbers of cells in any given study.
60 Extracellular recordings can overcome this limitation, but produce biased connectivtity
61 estimates, and therefore cannot be considered a gold standard.

62 Distinct imaging techniques provide distinct descriptors of neural processing, record-
63 ing at unique temporal and spatial resolutions. To take advantage of the synergy of in-
64 formation provided by multiple imaging approaches, we used a novel, large-scale, cross-
65 modal dataset: the Allen Institute's V1 Deep Dive (V1DD) ([1]). Following suit on the
66 proficiency of optical imaging techniques ([20]), V1DD provides multiple scans of high-
67 quality simultaneous two-photon calcium imaging (2PCI) recordings of thousands of
68 excitatory neurons within the mouse primary visual cortex (Fig. 1A). To record within
69 a cubic millimeter volume of V1 (800 μm x 800 μm x 800 μm), recording planes are sep-
70 arated by sixteen micrometers at thirty-seven frames per second as each plane is imaged
71 at six hertz. Identified regions of interest were run through a classifier trained on manual
72 labeling data meant to reduce false classification of non-soma ROIs as neuronal somas.
73 The 2PCI recording sessions were performed while the mouse intercepted various visual
74 stimuli, which included moving orientation gratings and natural movies. Along with its

75 exceptional functional recordings, V1DD also contains electron microscopy (EM) of the
76 same tissue volume, which has uncovered the fine-scale anatomy of the cubic millimeter
77 brain volume. EM has been employed extensively in large-scale datasets such as V1DD
78 to map the drosophila ([50]), worm ([61]), mouse ([18]), and even human brain ([53]).
79 The combination of these two imaging modalities has been applied in very few datasets
80 ([8, 33]), particularly at this scale. Four hundred cell-typed neurons in V1DD are co-
81 registered with verified correspondence between their fluorescence imaging ROI and a
82 morphologically typed reconstructed cell from the EM imaging. 315 of these correspond
83 to cells with activity traces recorded in the scan we analyzed (1-3). Out of these 315,
84 80 had fully reconstructed axons with all postsynaptic neurons identified, allowing for a
85 relation of the physiology of neural data to its exact anatomy at an unprecedented level.
86 The V1DD was explicitly chosen as it combines Ca^{2+} fluorescent recordings with high
87 spatiotemporal resolution of many fully reconstructed neurons in the EM data, proper-
88 ties that are essential to adequately accomplish the project’s goals. While the V1DD is
89 not yet publicly available or peer-reviewed, a comparable dataset was released by the
90 MICrONS Consortium ([12]) has been shown to be effective in other studies ([51, 57]).

91 With these advances offering new opportunities to examine the Hebbian cell as-
92 semblies at an unprecedented scale, we undertook and present in this work the largest
93 analysis to date of the relationship between structure and function within Hebbian cell
94 assemblies. Extracting assemblies from a Ca^{2+} fluorescence imaging scan, we examine
95 the reliability of their activation and their functional significance in the encoding of
96 visual stimuli. We then extended our analysis to the connectivity between neurons via
97 the cells which were coregistered with the structural EM scans. Deriving hypotheses di-
98 rectly from postulates advanced by Hebb, we tested his predictions about synaptic sizes
99 and connectivity both within and across assemblies, producing results which suggest a
100 far stronger role for inhibition in their formation and activation than is traditionally
101 assumed.

102 3 Results

103 3.1 Neuronal Organization of Hebbian Assemblies

104 SGC identifies neuronal assemblies using graph theory, treating assemblies as dis-
105 tinct graph communities. Unlike traditional methods that rely on pairwise correlations
106 between cells, SGC groups frames of fluorescence indicative of significant coactivity [44].
107 These moments of higher-order correlation are referred to as potential ‘phase sequences’,
108 representing when populations of neurons act cohesively as a closed circuit during assem-
109 bly activation [4]. Importantly, by design, SGC allowed for overlap between assemblies
110 and assigned a subset ($n = 748$) of neurons to no assemblies.

111 We analyzed a scan of the optical imaging dataset consisting of 2,708 excitatory
112 neurons recorded in parallel. The SGC algorithm generated 15 assemblies, which we
113 ordered by size with ‘A 1’ representing the largest assembly ($n = 1016$) and ‘A 15’
114 the smallest ($n = 23$). Their spatial distributions are seen in (Fig. 1C)*Will add 2D
115 xy plots with marginalized histograms. Although visually clear spatial organization is
116 not visible, statistical analysis using the Kolmogorov-Smirnov (KS) [39] test revealed
117 significant differences in spatial organization of the assemblies compared to that of the
118 entire neural space (Fig. 1E). In particular, there was marked organization in nearly
119 every assembly along the X-Y plane, indicative of the retinotopic organization observed
120 in studies of the visual cortex [56].

121 Assembly overlap has been integral to assembly studies in the past [32]. Hebb orig-
122 inally postulated that the shift between active assembly states could be what is col-
123 loquially referred to as a ‘train of thought’ [25]. Our results revealed a sizable degree

Extraction of Cell Assemblies in Large Scale Multi Modal Data

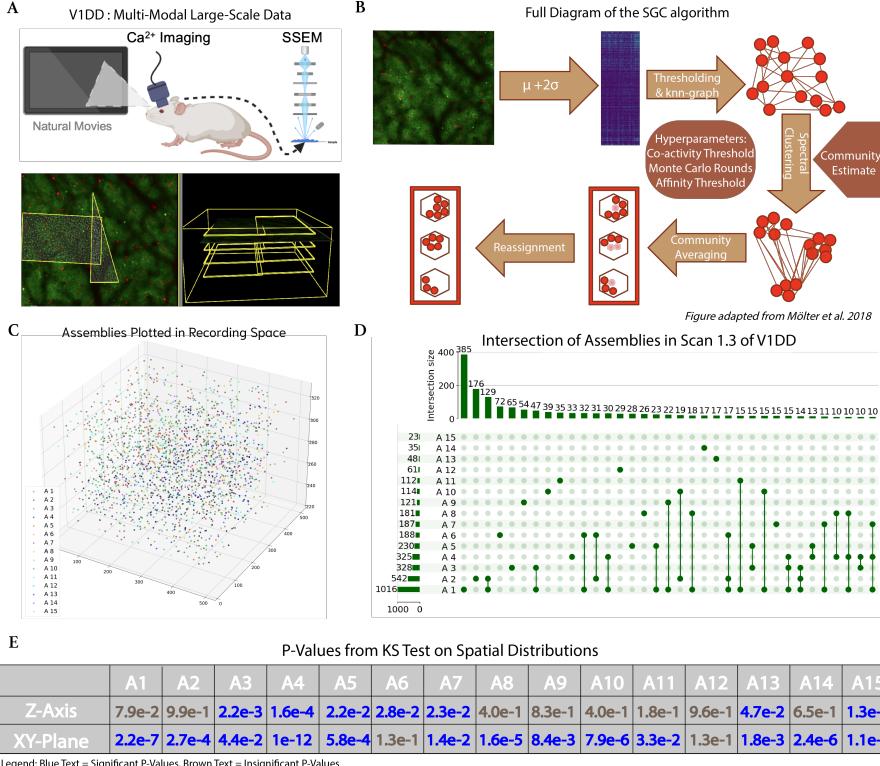


Figure 1: (A). V1DD consists of dense 2PCI recordings of thousands of excitatory neurons in a mm^3 section of mouse V1. Following, SSEM was leveraged to uncover the fine synaptic structure. The bottom left presents the calcium fluorescence of one scan while the bottom right presents the scan planes of the 2PCI microscope. As the V1DD is not yet publicly available, this panel was produced from a sister dataset, MICrONS. *To Do: Replace with an equivalent V1DD figure. Recording planes of V1DD are less spatially separated ($16 \mu\text{m}$). (B). SGC, an extraction algorithm uniquely designed for calcium imaging data, groups frames of the calcium fluorescence input to determine when neurons in assemblies are potentially coactivated. (C). Positions of fifteen extracted assemblies in the optical imaging recording space. Plot was produced by extracting the locations of all assigned cells in the recording field. In total, there are 1960 neurons plotted, including neurons assigned to multiple assemblies. The plot does not include 748 neurons that were assigned to no assemblies. Axes represent the spatial three-dimensions of the recording field, with units in micrometers. (D). An UpSet visualization of the intersection between assembly assignments. The histogram across the left represents the size of each individual assembly. The top histogram represents the size of the intersections between assemblies. Only intersections of ten neurons or greater were visualized. The center matrix indicates which assemblies belong to the subset formed by the top intersections. (E). A table presenting the results from the KS test on the spatial distributions of each neuronal assembly. Values colored in blue represent significant results (< 0.05), while brown signifies insignificant results. Assessment across the XY-Plane is meant to probe for retinotopic organization.

124 of overlap among assemblies, which was visualized through an UpSet plot [35], showing
125 neurons frequently assigned to multiple assemblies (Fig. 1D). The largest subset
126 ($n = 385$) consisted solely of neurons from 'A 1', while the most significant intersection
127 involved neurons shared by 'A 1' and 'A 2' ($n = 129$). Neurons from 'A 1' intersected
128 with ten other assemblies, highlighting that neural assemblies communicate with each
129 other through means greater than just functional interaction between synapses. They
130 possess an anatomical interaction through an extensive presence of overlapping neurons.

131 **3.2 Assemblies Reliability in Stimuli Response**

132 A section of the visual stimuli presented to the mouse consisted of natural movies.
133 Across the hour scan time, there were 15 unique 15-second movie clips. To evaluate
134 the functional reliability of neuronal assemblies, we analyzed their responses to these
135 stimuli. Reliability was quantified using an Oracle score, a leave-one-out correlation
136 metric (see methods) that measures the consistency of an activity trace across repeated
137 presentations of the same stimulus. This calculation was repeated for drifting gratings,
138 or stimuli where parallel bars move across the visual field. The neuronal activity during
139 the presentation of these gratings is used to characterize the reliability as a function of
140 orientation [19]. We sampled fourteen drifting grating presentations for each orientation.
141 Oracle scores have been used as a measure of the reliability of neuronal response [59].
142 To provide a baseline with which to compare the functionality of these assemblies to
143 tune to visual stimuli, the oracle scores of assemblies' coactivity traces were compared
144 to the oracle scores of all neuronal traces.

145 Assemblies exhibited significantly higher Oracle scores compared to the average re-
146 liability of individual neurons, indicating that assemblies, as populations, respond more
147 consistently to visual stimuli (Fig. 2A). To ensure this result was not merely due to
148 the inclusion of highly reliable neurons within assemblies, we separately calculated Or-
149 acle scores for neurons within assemblies and those assigned to no assemblies. Al-
150 though the comparison of the assemblies with all three cellular sets remained signifi-
151 cant (P-Values < 0.0001), there was no remarkable difference between the cellular sets
152 (P-Values > 0.25). This result suggests that the assembly's reliability stems from its
153 collective dynamics rather than the reliability of individual members. However, the
154 outliers of the cellular sets determined some highly reliable neurons that could perform
155 better than the Assembly trace.

156 Further analysis revealed consistent patterns of high assembly coactivity during spe-
157 cific visual frames of these natural movies. We defined 'trigger frames' as moments
158 when assembly activity exceeded a baseline threshold (see methods) and found that
159 these frames were highly consistent across repeated stimulus presentations. Example
160 trigger frame variance are presented in figure (Fig. 3B). Visualizations of these frames
161 suggest that assembly activity responds to complex features in the natural movies.

162 To further characterize their reliability, we computed Oracle scores of assembly
163 response to drifting gratings of various orientations (S. 9). Assemblies demonstrate
164 orientation-specific reliability as well, with response properties indicative of the under-
165 lying receptive fields of their individual neural components. These collective results
166 provide evidence of assemblies' ability to serve as functional populations with reliable
167 and specific responses to visual stimuli.

168 **3.3 Decoding Responses from Acute Visual Stimuli**

169 We also assessed the ability of assemblies to decode visual stimuli by implementing a
170 classification framework comparing assemblies to random ensembles. Random ensembles
171 served as a null model and were defined as randomly selected groups of neurons with

Assembly Reliability in Response to Acute Visual Stimuli

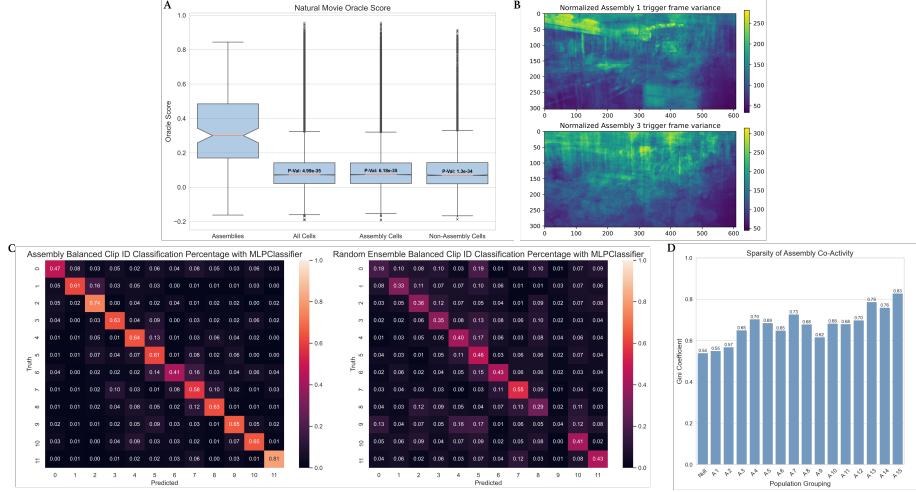


Figure 2: (A). Boxplot illustrating reliability of activity from assemblies and general neuronal populations in response to natural movies. Fifteen natural movies were repeatedly presented across the scan ten times. Oracle scores were used to exemplify reliability, with a leave-one-out mean of activity trace between repeats. Oracle scores of each assembly's coactivity trace were plotted, as well as the scores of all neurons ($n_{\text{all}} = 2708$). We also plot two other populations, neurons assigned to assemblies ($n_{\text{in}} = 1960$) and neurons that were not assigned to assemblies ($n_{\text{out}} = 748$). *These oracle scores are computed per stimulus, we plan to re-compute them by stimulus class, in order to properly account for sparsity and allow a comparison with randomly selected ensembles.(B). Plots of normalized ‘trigger frame’ variance of assemblies during natural movies. By averaging the frame of natural movies that assemblies were highly reactive to, in terms of peak co-activity, the natural movie frame was visually well reconstructed. Complementary results are seen in reliability of orientation gratings (S. 9). (C). Heatmap illustrating MLPClassifier decoding accuracy of natural movie clips with Assemblies and Random Ensembles. Heatmap values indicate accuracy of clip decoding by percentage of presentation. Clip IDs, indicating a unique natural movie clip, are balanced such that each clip had an equal frequency of presentation. Values in assembly heatmap are significantly greater than the random ensemble heatmap (Mann-Whitney u-stat: 6742.5 p-value: $2.89e^{-7}$, one-sided on diagonal elements u-stat: 139.0 p-value: $6.17e^{-5}$). (D). Histogram diagram of the sparsity (expressed as Gini Coefficient) of coactivity in cell assemblies and average sparsity of a null grouping of size-matched randomly selected neural ensembles. The null grouping’s coefficients were significantly smaller than the set of assembly coefficients (Wilcoxon Rank-Sum p=3.27e-5

172 the same size distribution as the assemblies. We employed a Multi-Layer Perceptron
173 Classifier (MLPClassifier) to evaluate how well each grouping could decode the identities
174 of the 15 natural movie clips. These classifiers have been shown to be effective in
175 academic and clinical settings, with high levels of accuracy and flexibility in available
176 hyperparameters [2] [7].

177 The results of our classifier revealed that assemblies significantly outperformed ran-
178 dom ensembles in accuracy. Heatmaps of classification accuracy for natural movie clip
179 identities demonstrated that assemblies provided more reliable decoding across repeated
180 trials. A Mann-Whitney U-test confirmed this finding, with assembly accuracy signifi-
181 cantly exceeding random ensembles for both overall performance (u-stat: 6742.5 p-value:
182 $2.89e^{-7}$) and diagonal elements (one-sided, u-stat: 139.0 p-value: $6.17e^{-5}$), indicating
183 enhanced stimulus-specific decoding.

184 To further characterize the functional properties of assemblies, we computed the
185 Gini coefficient [13] for each assembly’s activity trace. The Gini Coefficient, a statistical
186 measure of the ‘inequality’ of signal activity throughout the optical recording, revealed
187 that assemblies exhibited highly sparse activity patterns (Fig. 2D). The coefficient
188 for each assembly ranged from 0.55 to 0.83, with particular assemblies with extreme
189 sparsity, such as ’A 13’ (0.79) and ’A 15’ (0.83), exemplifying a high degree of functional
190 selectivity. Interestingly, this sparsity metric was not solely dependent on assembly size,
191 as intermediate-sized assemblies (e.g., ’A 4’ through ’A 12’) all exhibited similar Gini
192 coefficients of around 0.70.

193 The observed sparsity and decoding performance of assemblies reflect their role as
194 functional units optimized for efficient information processing. The selective activa-
195 tion patterns, characterized by high Gini coefficients, suggest that assemblies prioritize
196 encoding salient features of stimuli while minimizing redundant neural activity. This
197 coding strategy likely enhances the capacity for distinct representation of sensory in-
198 puts, aligning with the principle that coordinated activity in neuronal groups underpins
199 complex perception and cognition. In addition, the superior decoding accuracy of the
200 assemblies compared to random ensembles highlights their particular functional orga-
201 nization as fundamental processing modules. These findings reinforce the idea that
202 assemblies serve as coherent components of the nervous system, specialized for the in-
203 tegration and representation of sensory information.

204 **3.4 Structural Organization of Assemblies**

205 The availability of EM data enabled a unique opportunity to explore the structural
206 underpinnings of assemblies in the visual cortex. By mapping neural structure and
207 connectivity at a micrometer resolution [12] [57], we established an investigation of
208 the anatomical communication and organization of neurons with at least one shared
209 assembly membership (shared assembly cells) to those with disjoint membership. In
210 addition, we compared differences in connectivity of our assembly neurons to those
211 neurons not assigned to any assembly.

212 Our initial analysis of first-order connectivity surprisingly revealed no significant dif-
213 ferences in the probability of direct synaptic links between these sets or the strengths
214 of those links (Fig. 3C,F). This finding challenged assumptions that assemblies are
215 defined by densely interconnected excitatory neurons [25] and suggested that the defin-
216 ing structural characteristics of assemblies might lie beyond simple local connectivity
217 metrics.

218 To investigate higher-order structural patterns of these assemblies, we conducted a
219 motif analysis. In a neural network, when a significant number of subgraphs containing
220 a small number of interconnected cells repeat a particular pattern (e.g., Cell Type A →
221 Cell Type B → Cell Type A), they are classified as a motif [5] [54], with each subgraph

222 counted as a motif instance. The frequency with which these motifs persist has been
223 predictive of correlation in similar neural networks [28]. Motif analysis solidifies the
224 concept that synaptic plasticity promotes stable assembly formation by strengthening
225 connections within assemblies and weakening connections across assemblies.

226 Our analysis concentrated on second-order chain motifs, or structures consisting of
227 three neurons connected by two synaptic links (Fig. 3D). Second-order neural motifs
228 are divisible into various types based on the arrangement of their synaptic connections.
229 For this study, we prioritized chain motifs, as they align with synaptic plasticity rules
230 exhibiting an anti-symmetric profile. These rules, which are proposed to govern Hebbian
231 cells, create a ‘winner-take-all’ mechanism that preserves only one direction of a potential
232 bidirectional connection [16] [41]. To further refine our analysis, chain motifs were
233 categorized based on the type of intermediary neuron, distinguishing between excitatory
234 and inhibitory connections, with the latter providing insights into feed-forward inhibition
235 mechanisms.

236 Notably, while our excitatory chain analysis revealed no significant differences be-
237 tween shared and disjoint assembly memberships (Fig. 3D,G), an analysis of inhibitory
238 motifs convey a different story: disjoint assembly memberships exhibited considerably
239 stronger feed-forward inhibitory connections than shared memberships (Fig. 3E,H). This
240 suggests a mechanism of mutual inhibition that may regulate the interaction between
241 distinct assemblies, preventing excessive coactivation and ensuring modular processing
242 of neural information. The specificity of this finding underscores the unique role of
243 inhibitory motifs in shaping the dynamics and independence of neural assemblies, in-
244 dicative of a process missing from Hebb’s original theory.

245 3.5 Non-Assembly Cells

246 *Needs to be expanded to include other tests, such as insignificance in oracle scores
247 and insignificance in sparsity

248 In addition to motif analysis, we further assessed higher-order integration of assembly
249 neurons into the broader structural network using centrality metrics. In particular, we
250 found significantly lower betweenness centrality – a measure of a node’s importance in
251 mediating communication within a network – in neurons outside assemblies compared to
252 neurons within assemblies. This result indicates reduced participation in the structural
253 framework of the primary visual cortex, and imply a organization role of assembly
254 neurons as hubs for information flow. In contrast, non-assembly neurons may play a
255 more peripheral, secondary role in the network such as noise filtering.

256 All together, our structural results reveal assemblies are not anatomically defined
257 by a fundamental increase in local connectivity but through its higher-order patterns of
258 organization. The stronger feed-forward inhibitory connections in disjoint memberships
259 suggest a possible mechanism for maintaining functional segregation between assemblies,
260 while the enhanced network centrality emphasizes a role in neocortical information flow.
261 These structural insights provide for intuition on the mechanistic basis that drive the
262 functional properties of assemblies, reinforcing their contribution as modular units of
263 sensory processing.

264 4 Discussion

265 * We are planning on cutting the length of this section down in future revisions.
266 Suggestions on what is not the most important are highly welcome.

267 We leveraged a multimodal dataset from the mouse primary visual cortex, integrating
268 high-resolution calcium imaging to capture neuronal activity and EM to reveal struc-
269 tural connectivity. We identified distinct neuronal assemblies, each defined by unique

Structural Connectivity of V1 Neuronal Assemblies

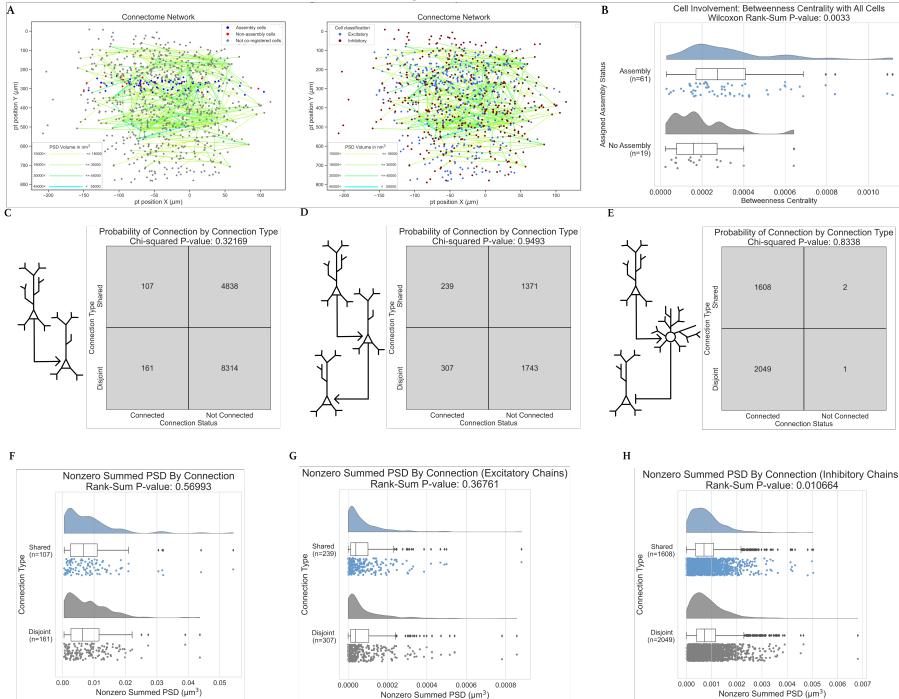


Figure 3: Statistical comparison of the organization in synaptic connectivity between coregistered cells assigned to assemblies and those not. All reconstructed neurons were found in layer 2/3 or layer 4. (A). The network being analyzed. (B) A raincloud plot of betweenness centrality, demonstrating a higher centrality for assembly neurons than those not in assemblies. (Wilcoxon Rank-Sum: P-Val < 0.05) (C-E). Chi-squared analysis of the likelihood of monosynaptic, disynaptic excitatory, and disynaptic inhibitory connections between neurons which share assemblies and neurons which participate in disjoint assemblies. (F-H) Raincloud plots showing the combined synaptic PSD volume per extant monosynaptic (F), disynaptic excitatory (G), and disynaptic inhibitory (H) connection, each divided between origin and terminus cell pairs which share assemblies and those which participate in disjoint assemblies.

270 coactivity patterns. These assemblies exhibited a spectrum of properties, including spatial
271 organization, overlap in membership, and acute response dynamics. In particular,
272 assemblies exhibited greater reliability to natural movies than the average neuronal trace
273 and significantly outperformed other ensembles in encoding the identity of natural visual
274 scenes. Complementing these functional insights, our structural analysis revealed that
275 while direct synaptic connectivity between the two sets was not considerably different,
276 there was markedly greater strength in feed-forward inhibition motifs involving cells of
277 disjoint membership than shared membership. Altogether, these results underscore the
278 utility of assemblies as cohesive units of sensory processing, and a mechanism for their
279 organization, bridging functional activity and structural organization.

280 Our analyses reveals three key organizational principles: First, assemblies exhibit
281 notable variability in size 1. SGC minimizes overestimation of the number of assemblies
282 or neuron assignments, suggesting that these size differences reflect intrinsic properties
283 rather than methodological artifacts. Larger assemblies, such as 'A 1', potentially serve
284 distinct roles compared to their smaller counterparts. Second, the extensive overlap of
285 neurons across multiple assemblies (Fig. 1D) supports the hypothesis of combinatorial
286 and hierarchical organization [66], with individual neurons contributing to multiple
287 functional modules. Third, while spatially overlapping, there is significant spatial orga-
288 nization of assemblies. Thus, formation of visual cortical assemblies may build upon
289 the same activity-dependent developmental processes recorded in cortical cells [46] [47]
290 [55]. Experience critical periods of development has been well demonstrated to cast and
291 whittle the arrangement of synaptic projections [34]. This refinement is well understood
292 to be critical for optimizing sensory processing, enhancing learning and memory forma-
293 tion [38], adaptability to changes in environments [17], skill learning [64], and cognitive
294 flexibility [58].

295 Assemblies demonstrated greater reliability in their response to naturalistic visual
296 stimuli compared to the average cellular response of all three sub-populations, as re-
297 flected by significantly higher Oracle scores. The decoding properties of assemblies
298 further validate their potential as representational modules in sensory processing. The
299 superior decoding capability of assemblies compared to random ensembles (Fig. 2C) em-
300 phasizes their ability to extract and represent key features of sensory inputs with high
301 fidelity. Potentially a key factor in the efficiency of their computation is the recorded
302 sparsity of assembly activation (Fig. 2D). In part, we consider the possibility that
303 this high degree of sparsity observed may have been indirectly imposed by SGC's noise
304 thresholding. However, we believe that the greatest influence is the inherent tempo-
305 ral dynamics of neural assemblies. The time scales at which assemblies operate in our
306 study, which match previous work, also coordinate relatively well with the time window
307 recorded for synaptic plasticity learning rules to operate *Will add discussion of Magee
308 and behavioral time scales The synchronization of cellular populations that have been
309 shown to propagate signals similar to the frequency of excitatory post-synaptic poten-
310 tials [22] provides evidence to validate the utility of assemblies in the visual system. In
311 the hippocampus, recorded signal sparsity has been proposed to be functional in encod-
312 ing and later retrieving memories [37]. However, the sparsity of ensemble coactivity in
313 the neocortex has yet to be well understood. Our working theory is that sparsity is a
314 manifestation of assembly cost-effectiveness. Sparse activation may ensure that assem-
315 blies focus neural resources on salient stimulus features while minimizing redundancy
316 and conserving computational capacity [45]. This sparsity may reflect a strategy in V1
317 computations to prioritize essential information, enabling the system to process complex
318 environments efficiently by capturing mutual information across neuronal components.

319 Furthermore, the collective reliability of assemblies may provide resilience against
320 representational drift. Representational drift reflects a functional characterization of the
321 fluctuations at the cellular level, such as synaptic turnover [27] [43]. This lack of stability

322 poses quite a challenge for attempting to predict the encoding properties of neurons,
323 which seem to fluctuate quite considerably over days [3] even if the stimulus environment
324 or behavior is stable over that period. Additionally, this phenomenon is not inherent
325 to the visual cortex. In the piriform cortex, the olfactory counterpart of the primary
326 visual cortex where odor identity is presumably encoded, there is a stark presence of
327 representational drift [52]. By maintaining consistent coactivity patterns, assemblies
328 offer a mechanism for preserving stable sensory representations despite cellular-level
329 variability.

330 In characterizing the structural underpinnings of neural assemblies, we offer a com-
331 pelling perspective on how inhibitory connections shape cortical organization. Although
332 Hebb's assembly theory was rooted in excitatory synapse modulation, our findings sug-
333 gest a crucial difference: the formation of Hebbian assemblies appears to rely on stronger
334 inhibitory indirect connections between assemblies, a mechanism which could be the re-
335 sult of anti-Hebbian plasticity. We argue, however, that this interpretation is consistent
336 with Hebb's overarching framework, which emphasized the plausibility of connectivity
337 models of perception, cognition, and behavior while deliberately constraining assump-
338 tions to mechanisms that were well-supported at the time. In doing so, Hebb laid a
339 foundation that accommodated future refinements and modern advances in our under-
340 standing of cortical organization.

341 These results also indicate that the plasticity mechanisms that generate these assem-
342 blies in developing networks are preferentially organizing connectivity in this chain-like
343 structure. This is in accordance with the theory of 'proto-ensembles' [66], which are
344 proposed to be the precursors to these population groupings. *Will add further detail
345 here. They are thought to be small, chain-like groups of neurons that are encouraged
346 to connect and form electrical synapses through their genetic code. Our results support
347 this theory, indicating there is a blueprint for cell assembly formation that integrates
348 the genetic, functional, and structural dynamics of the brain.

349 Several avenues for future research emerge from these findings. Integrating cell-
350 type-specific genetic information with functional assembly data could provide deeper
351 insights into the molecular foundations of assembly formation and maintenance [63].
352 Given that our optical imaging dataset focuses primarily on pan-excitatory neurons,
353 expanding this analysis to include inhibitory interneurons and other cell types could
354 offer a more comprehensive understanding of the functional and structural organization
355 of cortical cells. Furthermore, the dynamics of inter-assembly cooperation warrants
356 further investigation. Although our research provides evidence for the modularity of
357 assemblies, exploring how these units interact, particularly during phase sequences [11]
358 [36], could reveal mechanisms of large-scale neural coordination. This may uncover
359 hierarchical structures within and across assemblies, providing a stronger framework.
360 Finally, expanding the scope of this work to include downstream effects from higher-
361 order visual areas, such as the temporal cortex or hippocampus, could contextualize
362 these V1 assemblies within broader cortical and subcortical networks. Investigating
363 these interactions could bridge our understanding of local assembly dynamics with global
364 brain coordination, potentially advancing our knowledge of perception, memory, and
365 behavior.

366 5 Methods

367 5.1 Graph Clustering for Assembly Extraction

368 A paper by Mölter et al. compared eight prominent assembly extraction algorithms
369 upon application to both simulated and biological calcium imaging datasets, recording
370 performance as the simulated network size increased and the number of structurally

defined assemblies fluctuated [44]. Of the eight compared methods, their Similarity-Graph-Clustering (SGC) algorithm performed best on both datasets. While it did not yield a perfect reconstruction of the assemblies in the biological network, SGC could recover the assemblies with higher accuracy than all other algorithms. In part, this performance has been attributed to the computational effort SGC places in estimating the number of assemblies before defining them. For these reasons, we selected the SGC algorithm for assembly extraction from our dataset, and applied it With Dr. Molter's advice and assistance.

The SGC algorithm commences by thresholding the Calcium Fluorescence input to only signals that are two standard deviations above the mean to minimize noise (Fig. 1B). Afterwards, activity patterns are selected where the coactivity level of neurons exceeds the significance threshold. This threshold is a mutable parameter ($\alpha = 0.05$) that determines "high-activity patterns" based on a null model of all coactivity by shuffling the fluorescence signal (rounds = 1000). The network of high-activity activity patterns is then translated to a k -nearest-neighbors (kNN) graph where activity patterns are put into relation based on cosine difference. In this algorithm, k is not a hyperparameter and is chosen automatically to be the natural log of the number of nodes. The next step involves a community estimate, where a degree-corrected stochastic block model is fit to the kNN graph. The number of rounds this community estimate is averaged over is a hyperparameter that may be tuned (rounds = 5), as well as the statistical inference procedure (steps = 150000). With the estimate returning the most likely number of communities, spectral clustering applies this estimate to group the most similar activity patterns. These clusters are the first prospective selection of assemblies. However, the final step is a combination of community averaging and reassignment to reject groups that may have been erroneously defined because of a high level of noise in the original signal. This last step minimizes overestimating the number of assemblies or the neurons that should be assigned to those assemblies based on the affinity. This threshold for assigning neurons to assemblies, defined by the probability that the neuron was active in the assemblies' activity patterns, can be tuned as well (affinity = 0.4). The algorithm finally returns the extracted assemblies with assigned neurons. SGC allows for overlap between assemblies and assigns a part of the original data to no assemblies.

5.2 Oracle Scores

Oracle scores are a measure of the reliability of the trace response of cell activity to repeated visual stimuli computed through a jackknife mean, or leave-one-out mean [42] [49], of correlations between the activity trace across the repeated visual stimuli.

Let $S_i(t)$ denote the activity trace of a neuron or an assembly during the i -th presentation of a stimulus at time t . Let $\bar{S}_{-i}(t)$ represent the mean activity trace of all activity traces excluding the i -th presentation. From this, we can calculate the Oracle score O as:

$$O = \frac{1}{n} \sum_{i=1}^n \frac{\text{Cov}(S_i(t), \bar{S}_{-i}(t))}{\sqrt{\text{Var}(S_i(t)) \cdot \text{Var}(\bar{S}_{-i}(t))}}$$

In this formulation, the numerator will calculate the covariance between the activity trace at each presentation to the mean activity of all other repeats, while the denominator scales the magnitude of this covariance by the product of the standard deviations of $S_i(t)$ and $\bar{S}_{-i}(t)$.

414 **5.3 Trigger Frames**

415 To assess the visual stimuli associated with high activity in neuronal assemblies,
 416 we computed trigger frames by identifying peak activity times and extracting the corre-
 417 sponding images from a natural movie presentation. From a coactivity trace for assembly
 418 k over time $S_k(t)$, we can detect peaks P_k using the **scipy** signal package to define local
 419 maxima. We then define a mean trigger frame $\mu_k(x, y)$ for a pixel (x, y) as:

$$\frac{1}{|P_k|} \sum_{t \in P_k} I_t(x, y)$$

420 where $I_t(x, y)$ is the natural movie frame at time t . We then calculated the variance
 421 $\sigma_k^2(x, y)$ and normalized variance $\hat{\sigma}_k^2(x, y)$ for that pixel as:

$$\begin{aligned} \sigma_k^2(x, y) &= \frac{1}{|P_k|} \sum_{t \in P_k} (I_t(x, y) - \mu_k(x, y))^2, \\ \hat{\sigma}_k^2(x, y) &= \frac{\sigma_k^2(x, y)}{|P_k|} \end{aligned}$$

422 This normalized variance allowed us to visualize the trigger frame for each assembly.
 423 These computations allowed us to determine which frame of visual stimuli and consistent
 424 features in those stimuli were most strongly associated with high assembly coactivity.

425 **5.4 Decoder**

426 To evaluate the ability of assemblies to decode visual stimuli, we implemented a
 427 Multi-Layer Perceptron Classifier (MLPClassifier) from the **scikit-learn** library. This
 428 classifier was used to differentiate between 15 natural movie clips based on assembly
 429 coactivity time traces. Random ensembles of neurons with the same size distribution as
 430 the assemblies were used as a null model for comparison.

431 The MLP is a classical feedforward neural network composed of an input layer, one
 432 or more hidden layers, and an output layer. Each node in a layer is fully connected to
 433 every node in the subsequent layer through weighted connections. The final output of
 434 the function is determined by a non-linear activation function applied to the weighted
 435 sum of its inputs plus a bias term.

436 The MLPClassifier from **scikit-learn** is an implementation of a Multi-Layer Per-
 437 ceptron (MLP), a type of feedforward neural network. It operates by mapping input
 438 features \mathbf{x} to outputs $\hat{\mathbf{y}}$ through a series of hidden layers. Each layer consists of neurons
 439 that perform a weighted sum of their inputs, followed by the application of a non-linear
 440 activation function. Defining $\mathbf{W}^{(k)}$ as the weight matrix for the k -th layer, $\mathbf{h}^{(k)}$ as
 441 the input, and $\mathbf{b}^{(k)}$ as the bias vector for the k -th layer, we can define with activation
 442 function σ :

$$\mathbf{h}^{(k)} = \sigma \left(\mathbf{W}^{(k)} \mathbf{h}^{(k-1)} + \mathbf{b}^{(k)} \right),$$

443 For classification, the final layer uses the softmax activation function to output prob-
 444 abilities for each class. With L denoting the number of layers, we can define these output
 445 probabilities as:

$$\hat{\mathbf{y}} = \text{softmax}(\mathbf{h}^{(L)}) = \frac{\exp(\mathbf{h}^{(L)})}{\sum_j \exp(\mathbf{h}_j^{(L)})},$$

446 The MLPClassifier is trained using backpropagation, optimizing the weights and
447 biases via stochastic gradient descent or adaptive solvers such as Adam. Regularization
448 can be applied through an ℓ_2 -penalty term, controlled by a hyperparameter.

449 The assembly coactivity time traces were paired with corresponding natural movie
450 clip IDs. The data was split into training and test sets using an 80-20 split, and features
451 were scaled to normalize the input. A cross-validated grid search was used to optimize
452 the hyperparameters.

453 5.5 Gini coefficient

454 The Gini coefficient [13], a statistical measure that exemplifies the state of inequality
455 within a population. While often applied in economics to evaluate income inequality,
456 this metric has been applied as a valid approximation for signal sparsity [29] [31]. This
457 rendition of the application has been shown to serve as a relatively simple and robust
458 measure [62]. For our study, the coefficient is employed to quantify assembly signal
459 heterogeneity.

460 The Gini coefficient, G is often calculated with respect to the Lorenz curve, which
461 plots the cumulative distribution of a set (e.g., assembly coactivity trace) against its
462 rank in ascending order. For a given assembly A with coactivity trace $S = [s_1, s_2, \dots, s_n]$
463 where s_t is the proportion of active neurons at time point t , the coefficient for that
464 assembly is then calculated as:

$$G_A = \frac{\sum_{i=1}^{T-1} \sum_{j=i+1}^T |s_i - s_j|}{T^2 \cdot \bar{s}}$$

465 This computation is performed independently for each assembly, providing a metric
466 of signal inequality by normalizing with the total mean coactivity and the square of the
467 number of time points T . A G_A of 0 implies all values are identical, while a value of 1
468 indicates perfect inequality. A high G_A indicates that the coactivity is dominated by a
469 small number of time points, reflecting temporal sparsity.

470 5.6 Subgraph Enumeration

471 DotMotif [?] is an open-source tool used to find subgraphs and patterns within large-
472 scale graph networks. The algorithm detects subgraphs within a graph based on the
473 subgraph monomorphism principle.

474 For every graph $H = (N_1, E_1)$ that is given by the user, the algorithm detects
475 subgraphs G' within graph $G = (N_2, E_2)$ such that there exists a mapping $f : (N_1 \rightarrow N'_2)$
476 where $N'_2 \subseteq N_2$ and for every edge $\{a, b\} \in E_1$, the corresponding edge $\{f(a), f(b)\} \in$
477 E'_2 . The matched subgraphs may also contain additional edges within the graph G .
478 This algorithm was used to detect disynaptic chains within the connectome.

479 5.7 Motivating Postulates

480 Our statistical analyses involve tests of the following postulates. First, that excitatory
481 connections between cells which share at least one assembly ('shared' connections)
482 will be stronger than connections between cells which do not participate in any of the
483 same assemblies ('disjoint' connections), due to Hebbian plasticity. (Hebb, 1949, pp.)
484 This was examined both with regard to the post-synaptic density volume of monosy-
485 naptic connections between known coregistered cells within the dataset, and separately
486 in the form of the product of connection PSD volumes in disynaptic excitatory chains
487 which originated and terminated with shared or disjoint cells, allowing us to evaluate
488 indirect excitatory connections where the middle cell had not yet been coregistered.

489 Second, that excitatory connections will be more frequent within assemblies than be-
 490 tween assemblies, due to a combination of Hebbian plasticity and pruning of synapses.
 491 A number of computational studies have shown that long-term depotentiation (e.g. Het-
 492 herington I& Shapiro, 1993) and pruning (Garagnani, Wennekers I& Pulvermüller ,2009)
 493 play important roles in effective Hebbian assembly formation. This was examined in the
 494 form of per-connection targeting statistics, as well as per-cell inbound and outbound
 495 probability of both monosynaptic and disynaptic shared and disjoint connections.

496 Third, as discussed in the introduction |add section ζ , Hebb himself suggested that
 497 the assemblies could have been formulated via modulation of inhibition. Additionally, a
 498 number of computational models (e.g. Fransen, E., Lansner, A., Liljenström, H. 1993)
 499 have relied on inhibition between assemblies to restrict simultaneous activation, enabling
 500 competition between assemblies. It was therefore taken as an established hypothesis
 501 that one would expect inhibition between assemblies to be greater than within a given
 502 assembly. As monosynaptic connections between excitatory cells cannot be used to
 503 evaluate inhibition, this postulate was examined only in disynaptic chains, where the
 504 middle cell was morphologically classified as an inhibitory interneuron. Our examination
 505 involved the product of connection PSD volumes in such disynaptic inhibitory chains
 506 bridging shared and disjoint assembly neurons, along with a per-cell evaluation of the
 507 inbound and outbound probabilities of disynaptic inhibitory chain connections.

508 Fourth, and finally, we acknowledge that Hebb is discussing the reinforcement of
 509 sparse connectivity in his accounts of the emergence of cell assemblies (Hebb, 1959),
 510 and thus this reinforcement might have a significant effect on a sparse subset of connec-
 511 tions while producing a minimal difference in the central tendency of the overall set of
 512 connections. Dorkenwald et al. (2022) demonstrated a bimodality of the log PSD vol-
 513 ume of excitatory-excitatory synapses in the similar MICrONS mm^3 dataset, suggesting
 514 that a subset of such connections is impacted differently by processes determining PSD
 515 volume. Combining these two we decided to test the hypothesis that the larger of the
 516 two component distributions found in Dorkenwald et al. (2022) would be more likely
 517 than chance to involve connections between shared assembly neurons.

518 5.8 Statistical Methods

519 In this section, we detail the motivations and specifics of our analysis and methods.

520 We analyzed differences in connectivity metrics between shared-assembly and disjoint-
 521 assembly connection types, focusing on both the probability of connection and the
 522 strength of connections. Our analysis considered both direct monosynaptic and disynap-
 523 tic connections between neurons ("by connection"), as well as the sets of inbound and
 524 outbound connections grouped by cell ("by cell"). For disynaptic chains, we grouped
 525 by whether the intermediate (middle) cell was excitatory or inhibitory.

526 To evaluate connectivity metrics, we first defined the connection types based on
 527 assembly membership. Let A be the set of all assemblies, with A_j denoting the subset
 528 of assemblies that include pre-cell j , and A_i representing the subset of assemblies that
 529 include post-cell i . Formally,

$$A_j = \{a \in A \mid j \in a\} \quad \text{and} \quad A_i = \{a \in A \mid i \in a\}$$

530 Using these subsets, we defined the following binary indicators to capture the assem-
 531 bly relationship between pre-cell j and post-cell i :

- 532 • Shared _{ij} = 1 if $A_j \cap A_i \neq \emptyset$
- 533 • Disjoint _{ij} = 1 if $A_j \cap A_i = \emptyset$
- 534 • Shared_No_A _{ij} = 1 if $A_j, A_i = \emptyset$

- 535 • No_A_A_{ij} = 1 if $A_j = \emptyset$ and $A_i \neq \emptyset$
 536 • A_No_A_{ij} = 1 if $A_j \neq \emptyset$ and $A_i = \emptyset$

537 **5.8.1 Monosynaptic Metrics**

538 We defined w_{ij} as the pairwise summed post-synaptic density (PSD) between pre-
 539 cell j and post-cell i , and b_{ij} as an indicator variable that takes the value 1 if at least
 540 one synapse exists between j and i and 0 otherwise. All metrics exclude autapses, as
 541 these were not reliably represented in the dataset. ($j \neq i$).

542 The probability of monosynaptic outbound connection for a pre-cell j was calcu-
 543 lated as the proportion of realized connections under a given connection type $C \in$
 544 {Shared, Disjoint}, normalized by the total number of potential post-cell partners for
 545 that connection type:

$$b_{\text{out}_{j,C}} = \frac{\sum_{i|C_{ij}=1} b_{ij}}{|\{i \mid C_{ij} = 1, i \neq j\}|} \quad (1)$$

546 Similarly, the probability of monosynaptic inbound connection for a post-cell i was
 547 defined as:

$$b_{\text{in}_{i,C}} = \frac{\sum_{j|C_{ij}=1} b_{ij}}{|\{j \mid C_{ij} = 1, j \neq i\}|} \quad (2)$$

548 For connection strength, we computed the average realized summed monosynaptic
 549 outbound PSD for a pre-cell j as the total PSD across all post-cells satisfying the
 550 connection type C , normalized by the number of realized ($b_{ij} = 1$) connection under C :

$$w_{\text{out}_{j,C}} = \frac{\sum_{i|C_{ij}=1} w_{ij}}{\sum_{i|C_{ij}=1} b_{ij}} \quad (3)$$

551 Similarly, the average realized summed monosynaptic inbound PSD for a post-cell i
 552 is given by:

$$w_{\text{in}_{i,C}} = \frac{\sum_{j|C_{ij}=1} w_{ij}}{\sum_{j|C_{ij}=1} b_{ij}} \quad (4)$$

553 **5.8.2 Disynaptic Metrics**

554 In examining the inhibition in sets of cells that share assembly membership, and in
 555 sets that do not, we were primarily interested in describing inhibition being driven by
 556 the excitatory activity of an assembly's member cells. This moved us from the realm
 557 of monosynaptic connection analysis into an analysis of chains. Many aspects of our
 558 definition remained unaltered. A remained the set of all assemblies, with A_j the subset
 559 of assemblies that included pre-cell j , and A_i the subset of assemblies that included
 560 post-cell i . And the binary indicators indicating the assembly relationship between cells
 561 i and j remained unaltered.

562 But rather than simply using the monosynaptic weight between neurons j and i where
 563 $j \neq i$, we defined w_{ikj} as the product of the pairwise summed post-synaptic densities w
 564 in a three-cell chain motif with j as the first cell, an interneuron k as the second cell,
 565 and i as the third cell. Thus,

$$w_{ikj} = w_{ik} w_{kj} \quad (5)$$

566 And similar to our monosynaptic analysis, b_{ikj} was an indicator variable that took
 567 the value 1 if at least one disynaptic chain existed between j , k , and i and 0 otherwise.

568 Building on this definition, we outlined the method of normalization and metrics
 569 for disynaptic chain analysis, which accounted for both the intermediate and final (or
 570 first) cells in the chain. In this analysis, each pre-cell and post-cell was a coregistered
 571 excitatory neuron with an extended axon, consistent with the monosynaptic analysis.
 572 The intermediary cell in disynaptic chains, however, did not need to be coregistered or
 573 possess an extended axon.

574 Let n_e represent the number of excitatory cells and n_i represent the number of
 575 inhibitory cells from the set of all cells in the all-all connectome. Define $|k|$ be the
 576 number of potential middle partners. For inhibitory chains, $|k| = n_e$ as the middle cell
 577 is inhibitory. For excitatory chains, the middle cell cannot be the first or final cell, so
 578 $|k| = n_e - 2$.

579 The probability of disynaptic outbound connection for a pre-cell j was calculated
 580 as the proportion of realized disynaptic connections under a given connection type $C \in$
 581 $\{\text{Shared}, \text{Disjoint}\}$, normalized by the total number of potential chains satisfying the
 582 connection type:

$$b_{\text{out}_{j,C}} = \frac{\sum_k \sum_{i|C_{ij}=1} b_{ikj}}{|k| |\{i | C_{ij} = 1, i \neq j\}|} \quad (6)$$

583 Similarly, the probability of disynaptic inbound connection for a post-cell i was
 584 defined as:

$$b_{\text{in}_{i,C}} = \frac{\sum_k \sum_{j|C_{ij}=1} b_{ikj}}{|k| |\{j | C_{ij} = 1, j \neq i\}|} \quad (7)$$

585 For nonzero strength of connection, we computed the average realized summed disy-
 586 naptic outbound PSD for a pre-cell j as the total PSD across all chains satisfying the
 587 connection type C , normalized by the number of realized ($b_{ikj} = 1$) chains under C :

$$w_{\text{out}_{j,C}} = \frac{\sum_k \sum_{i|C_{ij}=1} w_{ikj}}{\sum_k \sum_{i|C_{ij}=1} b_{ikj}} \quad (8)$$

588 Similarly, the average realized summed disynaptic inbound PSD for a post-cell i was
 589 given by:

$$w_{\text{in}_{i,C}} = \frac{\sum_k \sum_{j|C_{ij}=1} w_{ikj}}{\sum_k \sum_{j|C_{ij}=1} b_{ikj}} \quad (9)$$

590 To facilitate statistical testing, we defined collections of metrics based on connection
 591 type C for both monosynaptic and disynaptic analyses. For disynaptic sets, the indices
 592 ij are replaced with ijk appropriately.

593 **Set Definitions**

594 The following sets were defined to evaluate connectivity metrics:

- 595 • The set of nonzero pairwise connection strengths under a given connection type
596 C :

$$\{w_{ij} \mid C_{ij} = 1\}.$$

- 597 • The set of outbound probabilities of connection for each pre-cell under a given
598 connection type C :

$$\{b_{\text{out}_{j,C}}\}.$$

599 Similar collections were defined for inbound probability of connection, and for
600 inbound and outbound nonzero average connection strengths by replacing the
601 metric accordingly.

- 602 • For inbound or outbound metrics where pre- or post-cells appear in both the
603 Shared and Disjoint groups, paired sets were constructed as:

$$\{b_{\text{out}_{j,C}}\}_{\text{paired}} = \{b_{\text{out}_{j,C}} \mid j \in \{j \mid b_{\text{out}_{j,\text{Shared}}} \} \cap \{j \mid b_{\text{out}_{j,\text{Disjoint}}} \}\}.$$

604 Similar paired collections were defined for inbound probabilities and both inbound
605 and outbound nonzero average connection strengths.

606 These sets provided the basis for the statistical tests used to compare metrics across
607 connection types.

608 **Statistical Tests**

609 We performed one-way statistical tests at $\alpha = 0.05$ to compare the Shared and
610 Disjoint groups. All alternative hypotheses predict Shared > Disjoint, except for tests
611 involving di-synaptic Inhibitory chain sets, in which the alternative hypotheses predict
612 Shared < Disjoint. We ran the following tests:

- 613 • For unpaired sets, we use a one-sided Wilcoxon Rank-Sum test
614 • For paired sets, we use a one-sided paired Wilcoxon Signed-Rank test to compare
615 metrics within cells appearing in both groups.

616 For pairwise binary connectivity, we created a contingency table to compare the
617 frequencies of successful and failed connections across connection types:

Connection Type	Successful Connections ($b_{ij} = 1$)	Failed Connections ($b_{ij} = 0$)
Shared	$\sum_{\text{Shared}_{ij}=1} b_{ij}$	$\sum_{\text{Shared}_{ij}=1} (1 - b_{ij})$
Disjoint	$\sum_{\text{Disjoint}_{ij}=1} b_{ij}$	$\sum_{\text{Disjoint}_{ij}=1} (1 - b_{ij})$

618 Then, we performed a Chi-Squared Test of Independence at $\alpha = 0.05$ to determine
619 if pairwise connection frequency differs across connection types.

620 **5.8.3 Tail Analysis**

621 In addition to the mono-synaptic and di-synaptic analyses, we performed a “tail”
622 (or “mode 2”) analysis to investigate whether the proportion of Shared versus Disjoint
623 connections differs between all pairwise connections and those classified as “tail” con-
624 nections.

625 To identify "tail" connections, we modeled the distribution of connection strengths
 626 using a Gaussian Mixture Model (GMM) with $k = 2$ components. The model was ini-
 627 tialized via k-means clustering to estimate the weights, means, and standard deviations
 628 of each component. The decision boundary separating the two Gaussian components
 629 was calculated as the intersection of their weighted probability density functions, derived
 630 using a quadratic equation based on the GMM parameters. Connections with values
 631 greater than or equal to the decision boundary were classified as "tail" connections. We
 632 present the model fit and evaluation as well as the tail boundary in the supplemental
 633 figures section.

634 Once the tail connections were identified, we compared the proportions of Shared
 635 and Disjoint connections in this subset to their proportions in the full dataset using a
 636 Chi-Squared Goodness-of-Fit Test at $\alpha = 0.05$. This test considered only the Shared
 637 and Disjoint groups, with expected proportions calculated relative to the total counts
 638 of these two groups in the full dataset.

639 5.8.4 Centrality analysis

640 Centrality analysis was used to quantify whether in a given network, assembly cells
 641 were more likely to be central to the network than non-assembly cells. This analysis
 642 gave further insights into the role of assembly cells in higher-order connectivity. To
 643 do this, we measured different centrality metrics for assembly and non-assembly cells,
 644 namely, in-degree centrality, out-degree centrality, closeness centrality, and betweenness
 645 centrality.

646 In a graph, the centrality of a node refers to its tendency to connect and generally
 647 influence other nodes within the network [6]. We developed a directed graph $G = (N, E)$,
 648 using the binary connectome such that N represents all the cells in the connectome
 649 and E represents the binary, directed connections between all cell pairs. The total
 650 number of outbound synaptic connections are given by $\sum \deg^+(n)$ and inbound synaptic
 651 connections are given by $\sum \deg^-(n)$. Normalizing these connections, in-degree centrality
 652 was calculated as:

$$I = \frac{\sum \deg^-(n)}{(N - 1)} \quad (10)$$

653 and out-degree centrality was calculated as:

$$O = \frac{\sum \deg^+(n)}{(N - 1)} \quad (11)$$

654 where $n \in N$.

655 A monosynaptic connection, $e = (j, i) \in E$ between a pre-cell j and post-cell i , will
 656 have i as the head and j as the tail end of the connection, where j and $i \in N$. The
 657 path between a pre-cell j acting as a source neuron and a post-cell i acting as a target
 658 neuron, is the alternating sequence of cells and connections starting from j and ending
 659 at i , with each cell before a connection in the sequence being a pre-cell and each cell
 660 after a connection being a post-cell. The number of monosynaptic connections within
 661 the path indicates the length of the path. The shortest path between cells j and i is the
 662 minimum length between the two cells. The shortest path, σ_{ij} , is also referred to as
 663 the geodesic path.

664 Based on this, the closeness centrality for a given cell i was calculated as the reciprocal
 665 of the summation of the shortest paths, or distances between the post-cell i and all other
 666 pre-cells j in the graph.

$$V(i) = \frac{N - 1}{\sum_{j=1}^{N-1} \sigma_{ij}} \quad (12)$$

667 For a given cell k , the betweenness centrality [14] was calculated as:

$$B_i(k) = \sum_E \frac{\sigma(k)_{ij}}{\sigma_{ij}} \quad (13)$$

668 where $\sigma(k)_{ij}$ is the number of shortest paths between pre-cell j and post-cell i that
669 pass through cell k . This value was also normalized to fall between $[0, 1]$.

670 **6 Code Availability**

671 All code is available at this GitHub link.
672 https://github.com/Jawagnercarena/Assembly_Organization
673 <https://github.com/AllenInstitute/HebbsVision>

674 **7 Acknowledgements**

675 TBD

676 **8 List of Contributions:**

677 TBD

678 **References**

- 679 [1] Abbasi-Asl, Reza, Josh Larkin, Kevin Takasaki, Daniel Millman, Daniel J. Denman,
680 Jerome Lecoq, Anton Arkhipov, Nathan W. Gouwens, Jack Waters, Marc Takeno,
681 Nuno Maçarico da Costa, R Clay Reid, and Saskia E. J. de Vries. V1 Deep Dive
682 Project. (*In Preparation*), 2024.
- 683 [2] Ayman H. Abdel-aziem and Tamer H. M. Soliman. A Multi-Layer Perceptron
684 (MLP) Neural Networks for Stellar Classification: A Review of Methods and Re-
685 sults. *International Journal of Advances in Applied Computational Intelligence*,
686 Volume 3(Issue 2):29–37, August 2023. Publisher: American Scientific Publishing
687 Group (ASPG).
- 688 [3] Kyle Aitken, Marina Garrett, Shawn Olsen, and Stefan Mihalas. The geometry
689 of representational drift in natural and artificial neural networks. *PLOS Compu-
690 tational Biology*, 18(11):e1010716, November 2022. Publisher: Public Library of
691 Science.
- 692 [4] Daniel G. Almeida-Filho, Vitor Lopes-dos Santos, Nivaldo A. P. Vasconcelos, José
693 G. V. Miranda, Adriano B. L. Tort, and Sidarta Ribeiro. An investigation of
694 Hebbian phase sequences as assembly graphs. *Frontiers in Neural Circuits*, 8:34,
695 April 2014.
- 696 [5] Uri Alon. Network motifs: theory and experimental approaches. *Nature Re-
697 views Genetics*, 8(6):450–461, June 2007. Number: 6 Publisher: Nature Publishing
698 Group.
- 699 [6] Alex Bavelas. A Mathematical Model for Group Structures. *Human Organization*,
700 7(3):16–30, March 2009.
- 701 [7] Amol Prataprao Bhatkar and G.U. Kharat. Detection of Diabetic Retinopathy in
702 Retinal Images Using MLP Classifier. In *2015 IEEE International Symposium on
703 Nanoelectronic and Information Systems*, pages 331–335, December 2015.
- 704 [8] Davi D. Bock, Wei-Chung Allen Lee, Aaron M. Kerlin, Mark L. Andermann, Greg
705 Hood, Arthur W. Wetzel, Sergey Yurgenson, Edward R. Soucy, Hyon Suk Kim, and
706 R. Clay Reid. Network anatomy and in vivo physiology of visual cortical neurons.
707 *Nature*, 471(7337):177–182, March 2011.
- 708 [9] L. G. Brock, J. S. Coombs, and J. C. Eccles. The recording of potentials from mo-
709 toneurones with an intracellular electrode. *The Journal of Physiology*, 117(4):431–
710 460, August 1952.
- 711 [10] György Buzsáki. Neural Syntax: Cell Assemblies, Synapsembles, and Readers.
712 *Neuron*, 68(3):362–385, November 2010.
- 713 [11] Luis Carrillo-Reid, Jae-Eun Kang Miller, Jordan P. Hamm, Jesse Jackson, and
714 Rafael Yuste. Endogenous sequential cortical activity evoked by visual stimuli.
715 *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*,
716 35(23):8813–8828, June 2015.
- 717 [12] MICrONS Consortium, J. Alexander Bae, Mahaly Baptiste, Agnes L. Bodor, Der-
718 rick Brittain, JoAnn Buchanan, Daniel J. Bumbarger, Manuel A. Castro, Brendan
719 Celii, Erick Cobos, Forrest Collman, Nuno Maçarico da Costa, Sven Dorkenwald,
720 Leila Elabbady, Paul G. Fahey, Tim Fliss, Emmanouil Froudakis, Jay Gager, Clare
721 Gamlin, Akhilesh Halageri, James Hebditch, Zhen Jia, Chris Jordan, Daniel Kap-
722 ner, Nico Kemnitz, Sam Kinn, Selden Koolman, Kai Kuehner, Kisuk Lee, Kai Li,

- 723 Ran Lu, Thomas Macrina, Gayathri Mahalingam, Sarah McReynolds, Elanine Mi-
 724 randa, Eric Mitchell, Shanka Subhra Mondal, Merlin Moore, Shang Mu, Taliah
 725 Muhammad, Barak Nehoran, Oluwaseun Ogedengbe, Christos Papadopoulos, Ste-
 726 lios Papadopoulos, Saumil Patel, Xaq Pitkow, Sergiy Popovych, Anthony Ramos,
 727 R. Clay Reid, Jacob Reimer, Casey M. Schneider-Mizell, H. Sebastian Seung, Ben
 728 Silverman, William Silversmith, Amy Sterling, Fabian H. Sinz, Cameron L. Smith,
 729 Shelby Suckow, Zheng H. Tan, Andreas S. Tolias, Russel Torres, Nicholas L. Turner,
 730 Edgar Y. Walker, Tianyu Wang, Grace Williams, Sarah Williams, Kyle Willie, Ryan
 731 Willie, William Wong, Jingpeng Wu, Chris Xu, Runzhe Yang, Dimitri Yatsenko,
 732 Fei Ye, Wenjing Yin, and Szi-chieh Yu. Functional connectomics spanning multiple
 733 areas of mouse visual cortex, July 2021. Pages: 2021.07.28.454025 Section: New
 734 Results.
- 735 [13] Robert Dorfman. A Formula for the Gini Coefficient. *The Review of Economics*
 736 and Statistics, 61(1):146–149, 1979. Publisher: The MIT Press.
- 737 [14] Linton C. Freeman. A Set of Measures of Centrality Based on Betweenness. *Soc-
 738 ciometry*, 40(1):35–41, 1977. Publisher: [American Sociological Association, Sage
 739 Publications, Inc.].
- 740 [15] Yves Frégnac. Aplysia: Hebbian or not? *Trends in Neurosciences*, 9:410, 1986.
- 741 [16] Júlia V. Gallinaro, Nebojša Gašparović, and Stefan Rotter. Homeostatic control
 742 of synaptic rewiring in recurrent networks induces the formation of stable memory
 743 engrams. *PLOS Computational Biology*, 18(2):e1009836, February 2022. Publisher:
 744 Public Library of Science.
- 745 [17] Erica R. Glasper and Gretchen N. Neigh. Editorial: Experience-Dependent Neu-
 746 roplasticity Across the Lifespan: From Risk to Resilience. *Frontiers in Behavioral*
 747 *Neuroscience*, 12:335, January 2019.
- 748 [18] Anjali Gour, Kevin M. Boergens, Natalie Heike, Yunfeng Hua, Philip Laserstein,
 749 Kun Song, and Moritz Helmstaedter. Postnatal connectomic development of in-
 750 hibition in mouse barrel cortex. *Science (New York, N.Y.)*, 371(6528):eabb4534,
 751 January 2021.
- 752 [19] Marc Green. Contrast detection and direction discrimination of drifting gratings.
 753 *Vision Research*, 23(3):281–289, January 1983.
- 754 [20] Christine Grienberger and Arthur Konnerth. Imaging Calcium in Neurons. *Neuron*,
 755 73(5):862–885, March 2012. Publisher: Elsevier.
- 756 [21] Amiram Grinvald, Amos Arieli, Misha Tsodyks, and Tal Kenet. Neuronal
 757 assemblies: Single cortical neurons are obedient members
 758 of a huge orchestra. *Biopolymers*, 68(3):422–436, 2003. eprint:
 759 <https://onlinelibrary.wiley.com/doi/pdf/10.1002/bip.10273>.
- 760 [22] Kenneth D. Harris. Neural signatures of cell assembly organization. *Nature Reviews*
 761 *Neuroscience*, 6(5):399–407, May 2005. Number: 5 Publisher: Nature Publishing
 762 Group.
- 763 [23] Kenneth D. Harris, Jozsef Csicsvari, Hajime Hirase, George Dragoi, and
 764 György Buzsáki. Organization of cell assemblies in the hippocampus. *Nature*,
 765 424(6948):552–556, July 2003.

- 766 [24] D. O. Hebb. The semiautonomous process: Its nature and nurture. *American*
 767 *Psychologist*, 18(1):16–27, 1963. Place: US Publisher: American Psychological
 768 Association.
- 769 [25] D.O. Hebb. *The Organization of Behavior*. John Wiley & Sons, New York, 1949.
- 770 [26] Anthony Holtmaat and Pico Caroni. Functional and structural underpinnings of
 771 neuronal assembly formation in learning. *Nature Neuroscience*, 19(12):1553–1562,
 772 December 2016.
- 773 [27] Anthony Holtmaat and Karel Svoboda. Experience-dependent structural synaptic
 774 plasticity in the mammalian brain. *Nature Reviews Neuroscience*, 10(9):647–658,
 775 September 2009. Number: 9 Publisher: Nature Publishing Group.
- 776 [28] Yu Hu, James Trousdale, Krešimir Josić, and Eric Shea-Brown. Motif statistics and
 777 spike correlations in neuronal networks. *Journal of Statistical Mechanics: Theory*
 778 and Experiment, 2013(03):P03012, March 2013. Publisher: IOP Publishing and
 779 SISSA.
- 780 [29] Niall Hurley and Scott Rickard. Comparing Measures of Sparsity. *IEEE Transac-*
 781 *tions on Information Theory*, 55(10):4723–4741, October 2009. Conference Name:
 782 IEEE Transactions on Information Theory.
- 783 [30] Jan Kamiński, Shannon Sullivan, Jeffrey M. Chung, Ian B. Ross, Adam N. Mame-
 784 lak, and Ueli Rutishauser. Persistently active neurons in human medial frontal and
 785 medial temporal lobe support working memory. *Nature Neuroscience*, 20(4):590–
 786 601, April 2017.
- 787 [31] Mahdi Khosravy, Naoko Nitta, Neeraj Gupta, Nilesh Patel, and Noboru Babaguchi.
 788 Chapter 3 - A descriptive review to sparsity measures. In Mahdi Khosravy, Nilanjan
 789 Dey, and Carlos A. Duque, editors, *Compressive Sensing in Healthcare*, Advances
 790 in ubiquitous sensing applications for healthcare, pages 43–63. Academic Press,
 791 January 2020.
- 792 [32] Andrei Kursin, Dušan Húsek, and Roman Neruda. Faster Learning with Overlap-
 793 ping Neural Assemblies. In Stefanos D. Kollias, Andreas Stafylopatis, Włodzisław
 794 Duch, and Erkki Oja, editors, *Artificial Neural Networks – ICANN 2006*, Lecture
 795 Notes in Computer Science, pages 226–233, Berlin, Heidelberg, 2006. Springer.
- 796 [33] Wei-Chung Allen Lee, Vincent Bonin, Michael Reed, Brett J. Graham, Greg Hood,
 797 Katie Glattfelder, and R. Clay Reid. Anatomy and function of an excitatory net-
 798 work in the visual cortex. *Nature*, 532(7599):370–374, April 2016. Number: 7599
 799 Publisher: Nature Publishing Group.
- 800 [34] Christiaan N. Levelt and Mark Hübener. Critical-Period Plasticity in the Vi-
 801 sual Cortex. *Annual Review of Neuroscience*, 35(1):309–330, 2012. _eprint:
 802 <https://doi.org/10.1146/annurev-neuro-061010-113813>.
- 803 [35] Alexander Lex, Nils Gehlenborg, Hendrik Strobelt, Romain Vuillemot, and
 804 Hanspeter Pfister. UpSet: Visualization of Intersecting Sets. *IEEE transactions on*
 805 *visualization and computer graphics*, 20(12):1983–1992, December 2014.
- 806 [36] Jason N. MacLean, Brendon O. Watson, Gloster B. Aaron, and Rafael Yuste. Inter-
 807 nal Dynamics Determine the Cortical Response to Thalamic Stimulation. *Neuron*,
 808 48(5):811–823, December 2005.

- 809 [37] Arnaud Malvache, Susanne Reichinnek, Vincent Villette, Caroline Haimerl, and
810 Rosa Cossart. Awake hippocampal reactivations project onto orthogonal neuronal
811 assemblies. *Science (New York, N.Y.)*, 353(6305):1280–1283, September 2016.
- 812 [38] S. J. Martin, P. D. Grimwood, and R. G. Morris. Synaptic plasticity and memory:
813 an evaluation of the hypothesis. *Annual Review of Neuroscience*, 23:649–711, 2000.
- 814 [39] Frank J. Massey. The Kolmogorov-Smirnov Test for Goodness of Fit. *Journal of*
815 *the American Statistical Association*, 46(253):68–78, 1951. Publisher: [American
816 Statistical Association, Taylor & Francis, Ltd.].
- 817 [40] Marcelo Matheus Gauy, Johannes Lengler, Hafsteinn Einarsson, Florian Meier,
818 Felix Weissenberger, Mehmet Fatih Yanik, and Angelika Steger. A Hippocampal
819 Model for Behavioral Time Acquisition and Fast Bidirectional Replay of Spatio-
820 Temporal Memory Sequences. *Frontiers in Neuroscience*, 12:961, December 2018.
- 821 [41] Christoph Miehl, Sebastian Onasch, Dylan Festa, and Julijana Gjorgjieva.
822 Formation and computational implications of assemblies in neural cir-
823 cuits. *The Journal of Physiology*, 601(15):3071–3090, 2023. eprint:
824 <https://onlinelibrary.wiley.com/doi/pdf/10.1113/JP282750>.
- 825 [42] Rupert G. Miller. A Trustworthy Jackknife. *The Annals of Mathematical Statistics*,
826 35(4):1594–1605, 1964. Publisher: Institute of Mathematical Statistics.
- 827 [43] Gianluigi Mongillo, Simon Rumpel, and Yonatan Loewenstein. Intrinsic volatility
828 of synaptic connections — a challenge to the synaptic trace theory of memory.
829 *Current Opinion in Neurobiology*, 46:7–13, October 2017.
- 830 [44] Jan Möller, Lilach Avitan, and Geoffrey J. Goodhill. Detecting neural assemblies
831 in calcium imaging data. *BMC biology*, 16(1):143, November 2018.
- 832 [45] Bruno A Olshausen and David J Field. Sparse coding of sensory inputs. *Current*
833 *Opinion in Neurobiology*, 14(4):481–487, August 2004.
- 834 [46] Janelle MP Pakan, Valerio Francioni, and Nathalie L Rochefort. Action and learning
835 shape the activity of neuronal circuits in the visual cortex. *Current Opinion in*
836 *Neurobiology*, 52:88–97, October 2018.
- 837 [47] Yuan Pan and Michelle Monje. Activity Shapes Neural Circuit Form and Function:
838 A Historical Perspective. *Journal of Neuroscience*, 40(5):944–954, January 2020.
839 Publisher: Society for Neuroscience Section: Viewpoints.
- 840 [48] Jesús Pérez-Ortega, Tzitzitlini Alejandre-García, and Rafael Yuste. Long-term
841 stability of cortical ensembles. *eLife*, 10:e64449, July 2021. Publisher: eLife Sciences
842 Publications, Ltd.
- 843 [49] Craig G. Richter, William H. Thompson, Conrado A. Bosman, and Pascal Fries. A
844 jackknife approach to quantifying single-trial correlation between covariance-based
845 metrics undefined on a single-trial basis. *NeuroImage*, 114:57–70, July 2015.
- 846 [50] Louis K Scheffer, C Shan Xu, Michal Januszewski, Zhiyuan Lu, Shin-ya Take-
847 mura, Kenneth J Hayworth, Gary B Huang, Kazunori Shinomiya, Jeremy Maitlin-
848 Shepard, Stuart Berg, Jody Clements, Philip M Hubbard, William T Katz, Lowell
849 Umayam, Ting Zhao, David Ackerman, Tim Blakely, John Bogovic, Tom Dolafi,
850 Dagmar Kainmueller, Takashi Kawase, Khaled A Khairy, Laramie Leavitt, Peter H
851 Li, Larry Lindsey, Nicole Neubarth, Donald J Olbris, Hideo Otsuna, Eric T Traut-
852 man, Masayoshi Ito, Alexander S Bates, Jens Goldammer, Tanya Wolff, Robert

- 853 Svirskas, Philipp Schlegel, Erika Neace, Christopher J Knecht, Chelsea X Al-
854 varado, Dennis A Bailey, Samantha Ballinger, Jolanta A Borycz, Brandon S Canino,
855 Natasha Cheatham, Michael Cook, Marisa Dreher, Octave Duclos, Bryon Eubanks,
856 Kelli Fairbanks, Samantha Finley, Nora Forknall, Audrey Francis, Gary Patrick
857 Hopkins, Emily M Joyce, SungJin Kim, Nicole A Kirk, Julie Kovalyak, Shirley A
858 Lauchie, Alanna Lohff, Charli Maldonado, Emily A Manley, Sari McLin, Car-
859 olinne Mooney, Miatta Ndama, Omotara Ogundeyi, Nneoma Okeoma, Christo-
860 pher Ordish, Nicholas Padilla, Christopher M Patrick, Tyler Paterson, Elliott E
861 Phillips, Emily M Phillips, Neha Rampally, Caitlin Ribeiro, Madelaine K Robert-
862 son, Jon Thomson Rymer, Sean M Ryan, Megan Sammons, Anne K Scott, Ash-
863 ley L Scott, Aya Shinomiya, Claire Smith, Kelsey Smith, Natalie L Smith, Mar-
864 garet A Sobeski, Alia Suleiman, Jackie Swift, Satoko Takemura, Iris Talebi, Dorota
865 Tarnogorska, Emily Tenshaw, Temour Tokhi, John J Walsh, Tansy Yang, Jane Anne
866 Horne, Feng Li, Ruchi Parekh, Patricia K Rivlin, Vivek Jayaraman, Marta Costa,
867 Gregory SXE Jefferis, Kei Ito, Stephan Saalfeld, Reed George, Ian A Meinertsha-
868 gen, Gerald M Rubin, Harald F Hess, Viren Jain, and Stephen M Plaza. A connec-
869 tome and analysis of the adult *Drosophila* central brain. *eLife*, 9:e57443, September
870 2020. Publisher: eLife Sciences Publications, Ltd.
- 871 [51] Casey M Schneider-Mizell, Agnes L Bodor, Forrest Collman, Derrick Brittain,
872 Adam Bleckert, Sven Dorkenwald, Nicholas L Turner, Thomas Macrina, Kisuk Lee,
873 Ran Lu, Jingpeng Wu, Jun Zhuang, Anirban Nandi, Brian Hu, JoAnn Buchanan,
874 Marc M Takeno, Russel Torres, Gayathri Mahalingam, Daniel J Bumbarger, Yang
875 Li, Thomas Chartrand, Nico Kemnitz, William M Silversmith, Dodam Ih, Jonathan
876 Zung, Aleksandar Zlateski, Ignacio Tartavull, Sergiy Popovych, William Wong,
877 Manuel Castro, Chris S Jordan, Emmanouil Froudarakis, Lynne Becker, Shelby
878 Suckow, Jacob Reimer, Andreas S Tolias, Costas A Anastassiou, H Sebastian Seung,
879 R Clay Reid, and Nuno Maçarico da Costa. Structure and function of axo-axonic
880 inhibition. *eLife*, 10:e73783, December 2021. Publisher: eLife Sciences Publications,
881 Ltd.
- 882 [52] Carl E. Schoonover, Sarah N. Ohashi, Richard Axel, and Andrew J. P. Fink. Rep-
883 resentational drift in primary olfactory cortex. *Nature*, 594(7864):541–546, June
884 2021. Number: 7864 Publisher: Nature Publishing Group.
- 885 [53] Alexander Shapson-Coe, Michał Januszewski, Daniel R. Berger, Art Pope, Yue-
886 long Wu, Tim Blakely, Richard L. Schalek, Peter Li, Shuhong Wang, Jeremy
887 Maitin-Shepard, Neha Karlupia, Sven Dorkenwald, Evelina Sjostedt, Laramie Leav-
888 itt, Dongil Lee, Luke Bailey, Angerica Fitzmaurice, Rohin Kar, Benjamin Field,
889 Hank Wu, Julian Wagner-Carena, David Aley, Joanna Lau, Zudi Lin, Donglai
890 Wei, Hanspeter Pfister, Adi Peleg, Viren Jain, and Jeff W. Lichtman. A connec-
891 tomic study of a petascale fragment of human cerebral cortex, May 2021. Pages:
892 2021.05.29.446289 Section: New Results.
- 893 [54] Olaf Sporns and Rolf Kötter. Motifs in Brain Networks. *PLoS Biology*, 2(11):e369,
894 November 2004.
- 895 [55] Gregory Z. Tau and Bradley S. Peterson. Normal Development of Brain Circuits.
896 *Neuropsychopharmacology*, 35(1):147–168, January 2010. Number: 1 Publisher:
897 Nature Publishing Group.
- 898 [56] Roger B. H. Tootell, Nouchine K. Hadjikhani, Wim Vanduffel, Arthur K. Liu,
899 Janine D. Mendola, Martin I. Sereno, and Anders M. Dale. Functional analysis
900 of primary visual cortex (V1) inhumans. *Proceedings of the National Academy of
901 Sciences of the United States of America*, 95(3):811–817, February 1998.

- 902 [57] Nicholas L. Turner, Thomas Macrina, J. Alexander Bae, Runzhe Yang, Alyssa M.
903 Wilson, Casey Schneider-Mizell, Kisuk Lee, Ran Lu, Jingpeng Wu, Agnes L. Bodor,
904 Adam A. Bleckert, Derrick Brittain, Emmanouil Froudarakis, Sven Dorkenwald,
905 Forrest Collman, Nico Kemnitz, Dodam Ih, William M. Silversmith, Jonathan
906 Zung, Aleksandar Zlateski, Ignacio Tartavull, Szi-chieh Yu, Sergiy Popovych, Shang
907 Mu, William Wong, Chris S. Jordan, Manuel Castro, JoAnn Buchanan, Daniel J.
908 Bumbarger, Marc Takeno, Russel Torres, Gayathri Mahalingam, Leila Elabbady,
909 Yang Li, Erick Cobos, Pengcheng Zhou, Shelby Suckow, Lynne Becker, Liam Panin-
910 ski, Franck Polleux, Jacob Reimer, Andreas S. Tolias, R. Clay Reid, Nuno Macarico
911 da Costa, and H. Sebastian Seung. Reconstruction of neocortex: Organelles, com-
912 partments, cells, circuits, and activity. *Cell*, 185(6):1082–1100.e24, March 2022.
- 913 [58] Lucina Q. Uddin. Cognitive and behavioural flexibility: neural mechanisms and
914 clinical considerations. *Nature Reviews Neuroscience*, 22(3):167–179, March 2021.
915 Number: 3 Publisher: Nature Publishing Group.
- 916 [59] Edgar Y. Walker, Fabian H. Sinz, Erick Cobos, Taliah Muhammad, Emmanouil
917 Froudarakis, Paul G. Fahey, Alexander S. Ecker, Jacob Reimer, Xaq Pitkow, and
918 Andreas S. Tolias. Inception loops discover what excites neurons most using deep
919 predictive models. *Nature Neuroscience*, 22(12):2060–2065, December 2019. Num-
920 ber: 12 Publisher: Nature Publishing Group.
- 921 [60] M. A. Wilson and B. L. McNaughton. Reactivation of hippocampal ensemble mem-
922 ories during sleep. *Science (New York, N.Y.)*, 265(5172):676–679, July 1994.
- 923 [61] Daniel Witvliet, Ben Mulcahy, James K. Mitchell, Yaron Meirovitch, Daniel R.
924 Berger, Yuelong Wu, Yufang Liu, Wan Xian Koh, Rajeev Parvathala, Douglas
925 Holmyard, Richard L. Schalek, Nir Shavit, Andrew D. Chisholm, Jeff W. Lichtman,
926 Aravinthan D. T. Samuel, and Mei Zhen. Connectomes across development reveal
927 principles of brain maturation. *Nature*, 596(7871):257–261, August 2021. Number:
928 7871 Publisher: Nature Publishing Group.
- 929 [62] Marina Wright Muelas, Farah Mughal, Steve O’Hagan, Philip J. Day, and Dou-
930 glas B. Kell. The role and robustness of the Gini coefficient as an unbiased tool
931 for the selection of Gini genes for normalising expression profiling data. *Scientific
932 Reports*, 9(1):17960, November 2019. Number: 1 Publisher: Nature Publishing
933 Group.
- 934 [63] Shengjin Xu, Hui Yang, Vilas Menon, Andrew L. Lemire, Lihua Wang, Fredrick E.
935 Henry, Srinivas C. Turaga, and Scott M. Sternson. Behavioral state coding by
936 molecularly defined paraventricular hypothalamic cell type ensembles. *Science*,
937 370(6514):eabb2494, October 2020. Publisher: American Association for the Ad-
938 vancement of Science.
- 939 [64] Atsushi Yamadori, Takashi Yoshida, Etsuro Mori, and Hikari Yamashita. Neu-
940 rological basis of skill learning. *Cognitive Brain Research*, 5(1):49–54, December
941 1996.
- 942 [65] R. Yuste, D. A. Nelson, W. W. Rubin, and L. C. Katz. Neuronal domains in
943 developing neocortex: mechanisms of coactivation. *Neuron*, 14(1):7–17, January
944 1995.
- 945 [66] Rafael Yuste, Rosa Cossart, and Emre Yaksi. Neuronal ensembles: Building blocks
946 of neural circuits. *Neuron*, pages S0896–6273(23)00967–4, January 2024.

9 Supplemental Figures

947

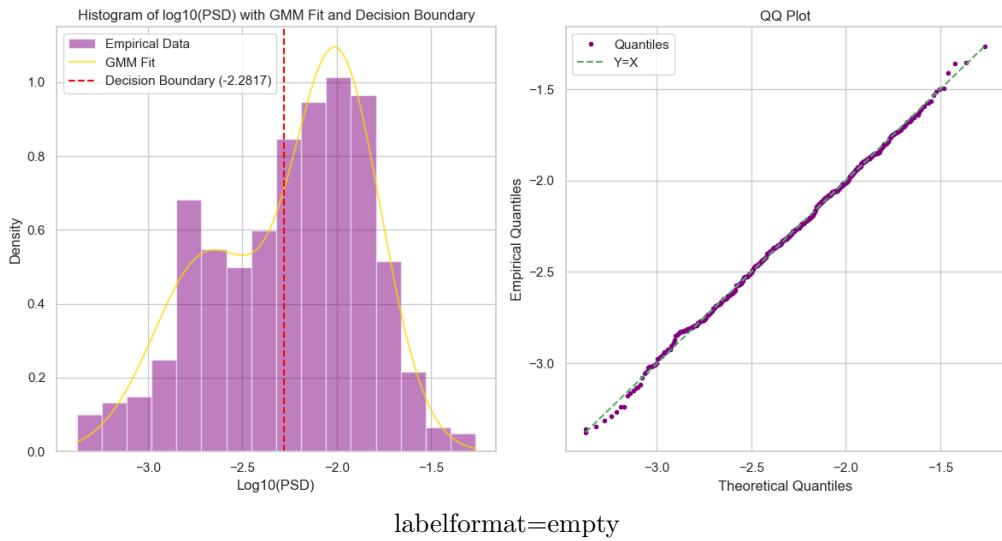
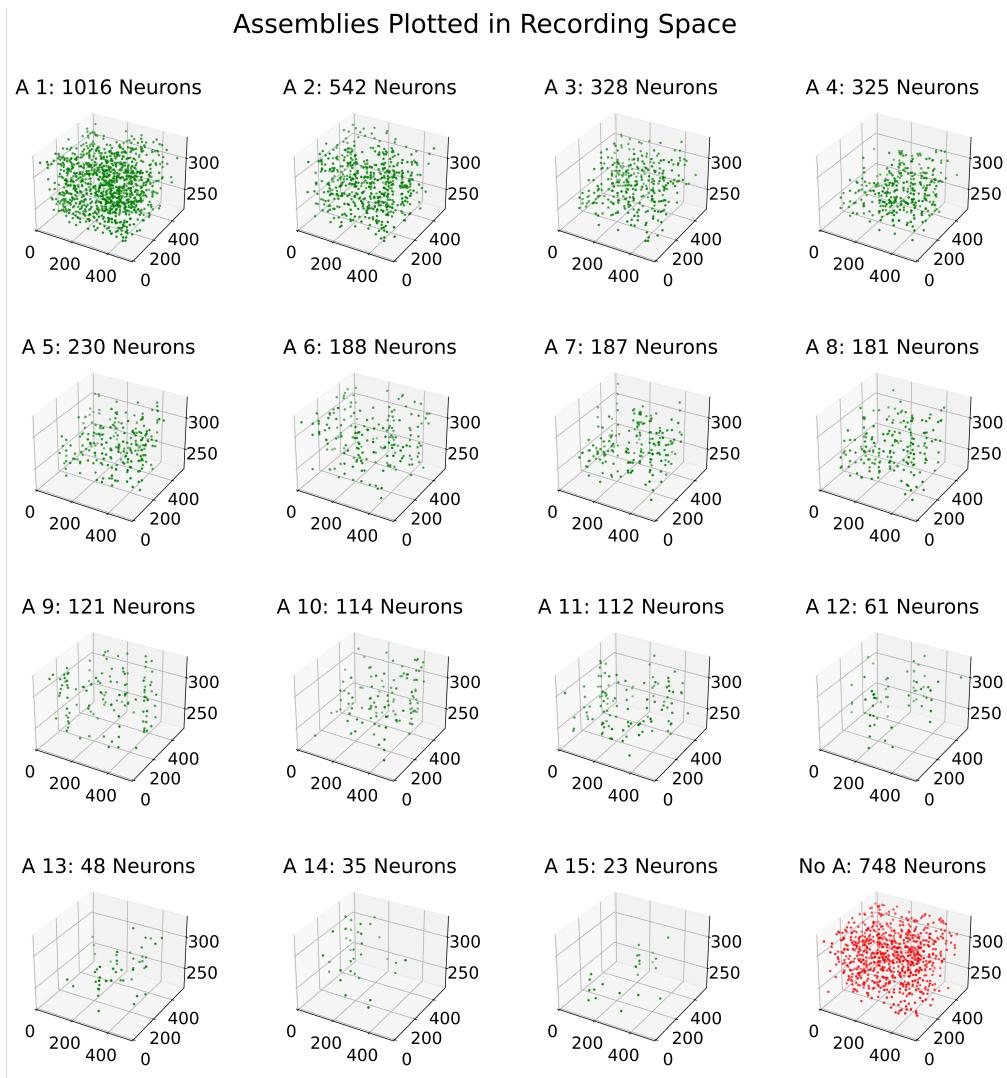
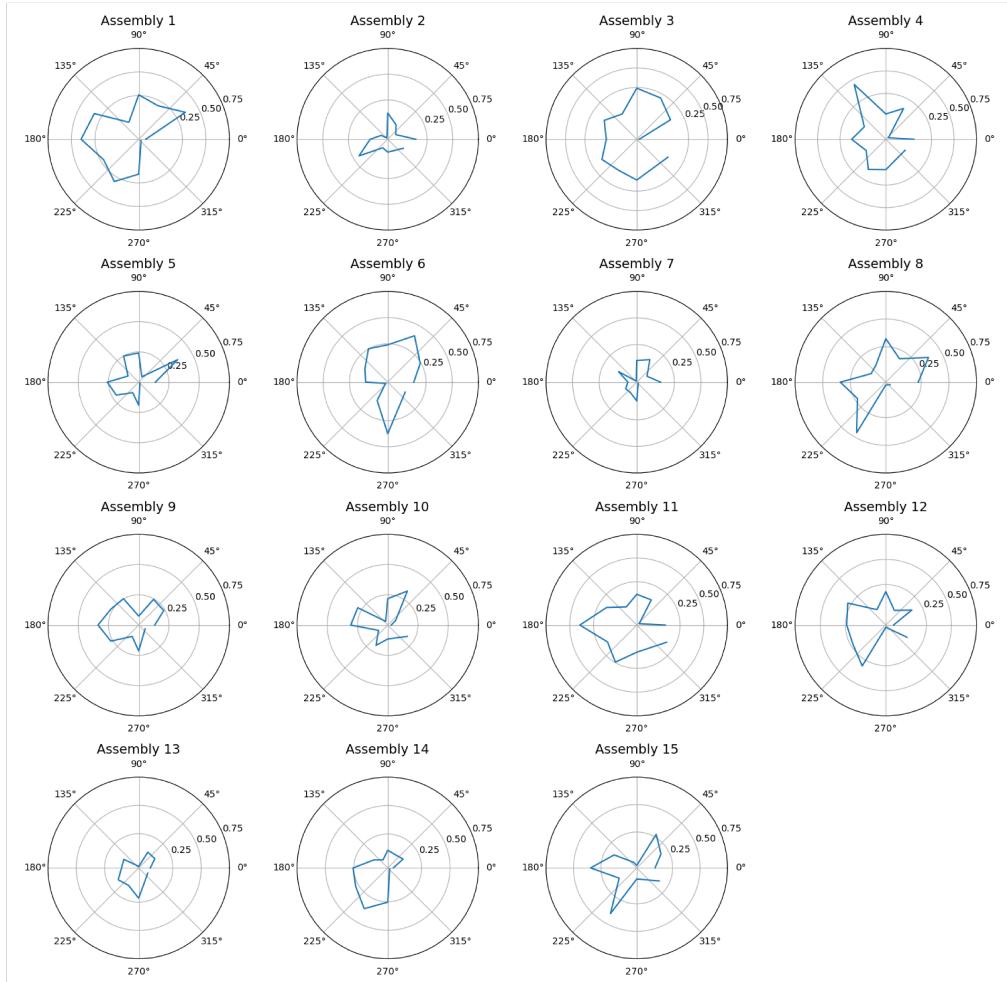


Figure 4: Supplementary figure 1. Histogram of connection strengths (log10-scaled) with GMM fit overlaid, and QQ plot evaluating the fit. The red dashed line indicates the decision boundary separating the two Gaussian components, used to classify "tail" connections.



Supplemental Figure 2. Individual Assemblies Plotted in 2PCI Recording Field. Every subplot provides an isolated view of an assembly, ordered by size, visualized in the optical imaging recording space. Each individual point refers to an identified excitatory neuron in the 2PCI. The plot of neurons assigned to no assemblies, 'No A', is also shown in red (bottom-right). Sub-plot axes refer to the three spatial dimensions of the recording field, with units in micrometers. Sub-plot titles also include the size of each assembly. In total, one thousand nine hundred and six neurons were assigned to assemblies and seven hundred forty-eight were not. Assemblies are typically spatially spread out throughout the recording field. This supplemental figure is meant to complement the visualization in Figure 5.

Assembly Oracle Score as a function of Fullscreen Grating Orientation



Supplemental Figure 3. Individual Assemblies Oracle Scores to FullScreen Gratings. Every sub-plot provides an isolated visualization of the reliability in an assembly's response with respect to the orientation of fullscreen gratings. Orientation of gratings are represented by a polar plot. Reliability is measured through the Oracle score metric. The scores of assembly co-activity trace in response to gratings are typically lower than those seen in natural movies (Fig. 8), but the results are still indicative of tuning properties in these functional populatinos which may be a product of the shared tuning of their individual neural components. Notably, some assemblies trace response seem to be highly reliable to particular orientations, similar to the orientation receptive fields of simple cells in the primary visual cortex.