Efficiency and reliability in biological neural network architectures

Daniela Egas Santander[1†*], Christoph Pokorny[1†], András Ecker[1], Jānis Lazovskis[2], Matteo Santoro[3], Jason P. Smith[4], Kathryn Hess[5], Ran Levi[6], and Michael W. Reimann[1]

[1]Blue Brain Project, École polytechnique fédérale de Lausanne (EPFL), Campus Biotech, 1202 Geneva, Switzerland
[2]Riga Business School, Riga Technical University, Riga, Latvia
[3]Scuola Internazionale Superiore di Studi Avanzati (SISSA), Trieste, Italy
[4]Department of Mathematics, Nottingham Trent University, Nottingham, UK
[5]UPHESS, BMI, École polytechnique fédérale de Lausanne (EPFL), 1015 Lausanne, Switzerland
[6]University of Aberdeen, Aberdeen, UK
[†]Co-lead authors
[*]Correspondence: daniela.egassantander@epfl.ch

Abstract

Neurons in a neural circuit have been demonstrated to have astonishing diversity in terms of numbers and targets of their synaptic connectivity and the statistics of their spiking activity. We hypothesize that this is the result of an underlying struggle between reliability, robustness and efficiency of the information represented by their spike trains. Specifically, certain architectures of connectivity foster highly uncorrelated and thus efficient activity, others foster the opposite trends, i.e., robust activity. Both coexists in a neural circuit, leading to the observed long-tailed and highly diverse distributions of connectivity and activity metrics, and allowing the robust subpopulations to promote the reliability of the network as a whole. To test the hypothesis and characterize these architectures, we analyzed several openly available connectomes and found that all of them contained groups of neurons with very different levels of complexity of their connectivity. Using co-registered functional data and simulations of a morphologically detailed network model, we found that low complexity groups were indeed characterized by efficient spiking activity and high complexity groups by reliable but inefficient activity. Moreover, for neurons in cortical input layers, the focus was increasing reliability; for output layers, it was increasing efficiency. To test the effect of the complex subpopulations on the reliability of the network as a whole, we manipulated the connectivity in the model to increase or decrease complexity and confirmed that it affected activity in the expected ways. Our results impact our understanding of the neural code, demonstrating that it is as diverse as neuronal connectivity and activity, and must be understood in the context of the efficiency/reliability struggle.

1 Introduction

There is a delicate balance that determines an optimally functioning network. If the activity of the neuronal population is highly correlated the information is encoded robustly. Indeed, the correlations add redundancy, thus providing a form of population code that can overcome noise sources such as synaptic failure[1–3] or loss of neurons. On the other hand, if the activity

is mostly uncorrelated the information is encoded efficiently. While the goals of *efficiency* and *robustness* are overall mutually exclusive, they do not have to be uniform within a group of neurons. Brain regions and microcircuits have been characterized as having distinct input and output populations[4,5] that could use different solutions to this struggle. Additionally, it has been demonstrated that correlated (thus robust) synaptic inputs lead to more reliable responses across trials on the single neuron level[6,7]. Given the high recurrence of biological neural networks, this indicates that robust populations promote *reliability*.

In biology, correlated spiking emerges in the form of neuronal assemblies[8–10], overlapping groups of neurons that fire together. Additionally, activity has been demonstrated to be constrained to a lower-dimensional manifold[11]. Both are indications of sacrificing efficiency for robustness. These functional features are thought to be implemented by the structure of the synaptic connectivity. In simulations, researchers created assemblies by modeling more or stronger connections within than across assemblies[12–15]. However, biological circuits have a non-random connectivity structure[16–18] that cannot be captured by this class of models. A way to demonstrate this quantitatively, is to consider the structure of the connectivity emerging from *higher-order* interactions, i.e., beyond the pairwise level, in the connectivity of larger groups of neurons[19,20].

We hypothesize that this non-random structure solves the struggle between efficiency and robustness, allowing a circuit to implement different solutions along a spectrum between the two mutually exclusive goals in different subnetworks. Additionally, it serves to convert the robustness in a redundant subpopulation code to reliability on the single neuron level across the whole network. This perspective provides an explanation for the well-established, long-tailed distributions of structural and spiking-related parameters: Structural diversity mirrors diversity of functional priorities in different subnetworks[21].

We analyzed various connectomes with cellular resolution, ranging from the worm to an electron-microscopic reconstruction of mouse cortical connectivity, to find topological measurements that characterize differences between subnetworks that are likely to be functionally relevant. We then used functional data that were co-registered to the electron microscopic MICrONS connectome[22] and simulations of a morphologically detailed model[23] to confirm the relevance of the metrics for reliability, robustness and efficiency, and develop a detailed picture of the structure - function relation.

Intuitively, the high correlations within functional assemblies or other robust subnetworks could be achieved by more heavily tying together their subpopulation. This structure could be measured either through pairwise interactions, i.e., connection probability or overexpression of reciprocal connections, or through the higher-order structure usually described in terms of rich clubs, clustering or small world coefficients Towlson et al.[18], Shih et al.[24], Turner et al.[25], Manjunatha et al.[26]. Here, we employ a different mathematical framework. It is based on finding densely connected motifs called *directed simplices* and describing how they are connected to each other to form larger motifs. A directed simplex of dimension d is a motif of $d+1$ neurons where the connectivity between them fulfills the following criterion: There exists a numbering of neurons in the motif from 0 to d, such that if $i < j$ then there exists a synaptic connection from n_i to n_j (Fig. 1A). We also call neuron n_0 the *source* and neuron n_d the *sink* of the simplex. This framework offers great analytical flexibility as it defines a range of network metrics that describe not only the global structure of a graph, but also local, i.e., node- or even edge-based metrics. Additionally, simplices have been demonstrated to have functional relevance. In particular, they are related to the reduction of dimensionality of neural activity and the formation of assemblies, as the correlation of spiking activity of neurons increases with the dimension of simplices they participate in[19,27]. Finally, while simplices have been shown to be over-expressed motifs in neural networks at all scales[28–31], constructing random sparse directed graphs with a

given overexpression of simplices is an unsolved mathematical problem[32,33]. This demonstrates that these motifs are relevant, significant and non-trivial.

We confirmed the overexpression of directed simplices in all connectomes we studied, and found that reciprocal connections are more abundant in these motifs, especially for high dimensions. We developed a metric for the local complexity of subnetworks centered around individual neurons (a neuron's *neighborhood*), which characterizes how much it differs from a random subnetwork of the same size. Notably, we demonstrate that all connectomes have an unexpected amount of structural diversity in the neighborhoods that goes beyond the one characterized before (e.g. long-tailed degree distributions) and that our complexity metric captures this variability. Relating this finding to our initial hypothesis, we predict that the activity in complex neighborhoods is highly correlated and hence robust rather than efficient. We confirm this using activity data co-registered to the MICrONS connectome[22] and simulations of a morphologically detailed network model[23]. We further confirmed that spike trains of efficient neighborhoods provide better readouts in a simulated stimulus classification task, and that known cortical input layers seem to be optimized for increasing reliability and output layers for increasing efficiency. To demonstrate that this is causation and not just correlation, we simulated connectomes that were manipulated to increase or decrease their global complexity and found that reliability changed as predicted. Finally, we contrasted the impact of complex subnetworks on the global and local levels, i.e., whether changes in efficiency and reliability affect all neurons equally or a select few. We found that increases in reliability were global, while high efficiency was limited to a smaller number of neighborhoods that may serve as outputs of the local circuit. Our results show the importance of the rich internal structure that is revealed when connectivity is considered at cellular resolution, which must be considered to fully understand the network's function. They also show that parts of the complex network structure in brain circuitry may be dedicated to overcoming neuronal variability; hence sources of that variability, such as synaptic failure, must be considered to fully understand the structure-function relation.

2 Results

2.1 The overexpression of reciprocity is shaped by the higher-order structure

To confirm previous results about the structural relevance of reciprocal connectivity and simplex motifs, we counted their overexpression in the following four available connectomes at cellular resolution: First, the electron microscopy reconstructions of the full brain connectivities of the adult *C. elegans*[34] and the *Drosophila larva*[35]. Second, the electron microscopy reconstructions of $1\,mm^3$ volume of mouse primary visual cortex (*MICrONS*[22]) and a morphologically detailed model of $\sim 2\,mm^3$ of rat somatosensory cortex (*BBP*[23]). See Methods for details on the data sources for all connectomes.

Across all connectomes we found an overexpression of reciprocal connections and directed simplex counts when contrasted to relevant control models of the same size (Fig. 1 B, C). The controls considered were: *Erdős–Rényi* controls (ER), *directed configuration model* controls (CM) and *distance dependent* controls (Dist), which preserve: the overall density of connections, the in and out degrees of the nodes, and the density of connections at a given distance, respectively (see Methods). Note in particular, that the CM controls will preserve the known long-tailed degree distributions[21,36–38]. The distance dependent controls where only performed for MICrONS and BBP, given that Euclidean distances between cell bodies might not be the most relevant factor determining connection probability for C. elegans and Drosophila. Overexpression of reciprocal connections was weakest in the BBP model connectome, to a degree where it was

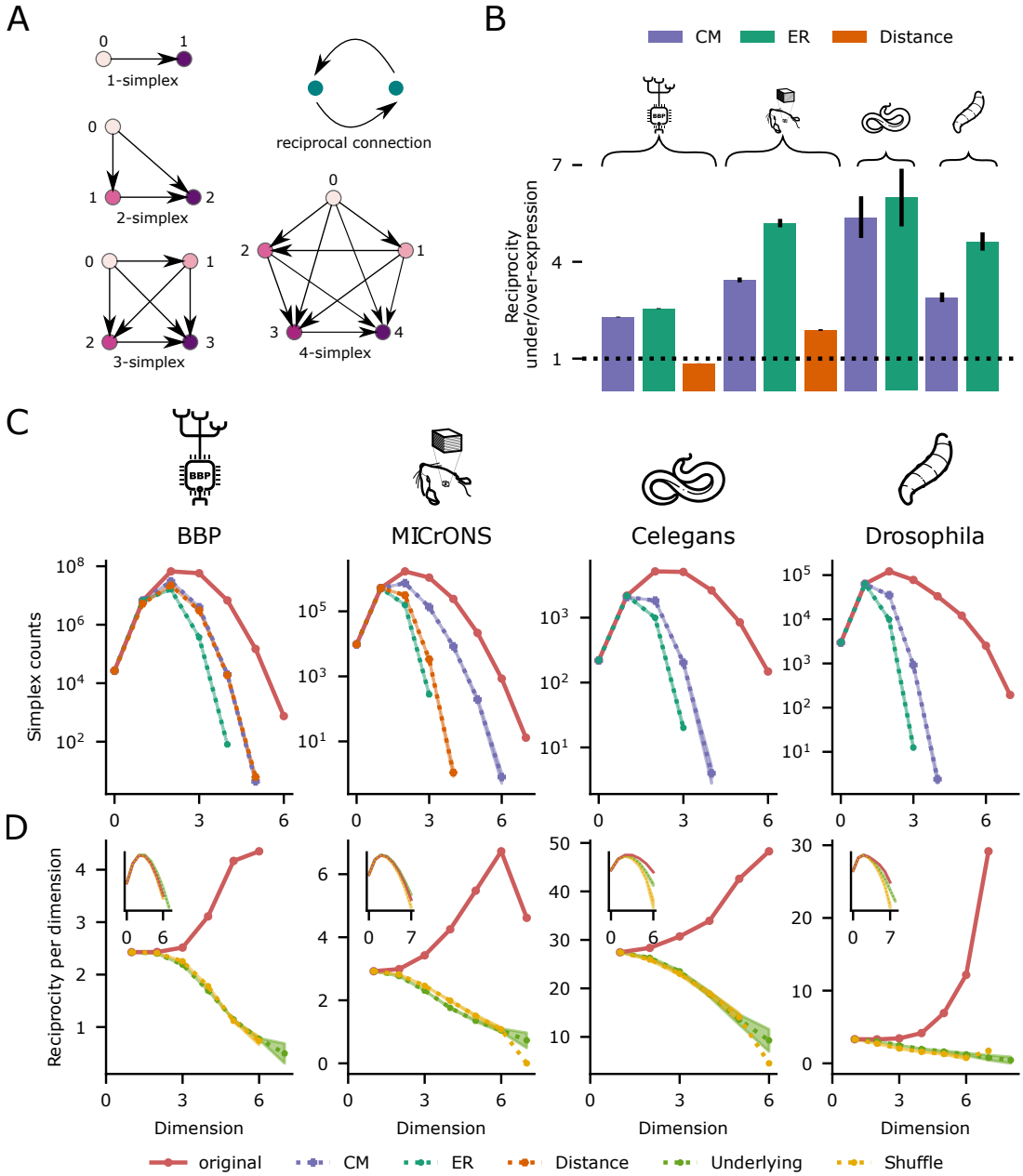

Figure 1: Over expression of directed motifs. A: Reciprocal connection (green nodes) and simplices (purple nodes). **B:** Percentage of reciprocal connections in each connectome relative to mean of the percentage of 100 controls of different types: Erdős–Rényi, Configuration model and Distance dependent (see Methods). A value of 1 indicates the original connectome has the same percentage of reciprocal connections than its corresponding control; values above/below 1 represent over/under expression; vertical bars show the standard deviation. All values are significantly different than the mean with p-values under 3×10^{-155} for a two sided one sample t-test. **C:** Overexpression of simplex motifs with respect to 10 random controls, which by design match the counts of the original network for dimensions 0 and 1. All counts in dimensions greater than one are significantly higher than the mean of the controls with p-values under 4×10^{-18} for a one sided one sample t-test. **D:** Percentage of reciprocal connections in the subgraph of simplices of each dimension contrasted with the same curve for controls where just the directionality of the connections is modified (see Methods). Inset: Simplex counts of the directional controls are close to the original ones.

not overexpressed compared to the distance-dependent control. All other overexpressions of reciprocal connections and simplex counts were statistically significant at the most stringent significance levels. In particular, this shows that distance dependent connectivity and degree heterogeneity contribute but do not fully explain the high complexity of the connectivity structure across species.

In addition to simplex counts as a global metric of network complexity[39–41], we can measure the participation of individual elements in that structure. For example, *k-edge participation*, defined as the number of k-simplices an edge is part of gives a notion of edge centrality that has been observed to be meaningful for describing synaptic strength, and plastic changes in it, both in MICrONS and the BBP connectome[42]. We can filter edges by the maximum dimension at which this metric is non-zero. Doing this, we found that edges with a high maximum dimension are more likely to be reciprocal and this preference increases with dimension (Fig. 1D, red curves). Moreover, this preference is not merely a combinatorial artifact. To show this, we contrasted these curves with those obtained from two types of random controls: one (*rc-Sh*) only modifies the location of reciprocal connections and the other (*Und*) modifies their location together with the direction of all the connections (see Methods). Crucially, these controls are quite stringent in that they have the same underlying undirected graph as the original and thus approximately preserve the number of simplices across dimensions (Fig. 1D, inset). We observed that the percentage of reciprocal connections in these controls actually decreases with dimension, showing that combinatorially it is less likely for a reciprocal connection to exist on higher dimensional maximal simplices if these where placed at random.

In summary, we confirmed previous results that reciprocal connections (a circular motif) and simplices (a feedforward motif) are strongly overexpressed in biological neural networks (BNNs); and that their locations within the networks are strongly related. Moreover, these metrics allowed us to show that the the complexity of BNNs can not be explained by only distance dependent connectivity or long tailed degree distributions.

2.2 Biological neural networks have a large diversity of neighborhoods

While in the previous section we have demonstrated a non-random global structure of BNNs, we were also interested in the diverse ways in which individual neurons participate in this structure, and how it affects their function. We can do this in two complementary ways: First, relating to the *simplicial structure*, by taking averages of functional or structural properties of neurons weighted by the number of maximal simplices they participate in (see Methods). Second, relating to their synaptic *neighborhood structure*, given by all neurons directly interacting synaptically with a central one. To do this, we define the *neighborhood* of a neuron v, as the subnetwork on v together with all neurons connected to it (in either direction) and all connections between them and we call v its *center* (Fig. 2A).

One can apply any global network metric to neighborhoods, a particularly simple example is the neighborhood *size* the number of nodes in it, which closely approximates the sum of the in and out degree of its center (see Methods). As expected from previous research[38,37,18,43], we found across connectomes, that the distribution of neighborhood sizes is has a longer tail than in its corresponding controls, except for CM controls which fix the node degrees and thus the neighborhood sizes. However, for a plethora of other network metrics, the distributions of their values are also more long tailed than all corresponding controls (see Fig. S1 for examples) including CM, indicating that this difference is not driven by long-tailed node degree distributions alone.

This suggest that there is a variety of complexity of the neighborhoods of a BNN, which according to our hypothesis (stated in the introduction) should fulfill different functional roles. To test this, we need to be able to capture this complexity in a single number which is independent

of the neighborhood size. We propose a metric of neighborhood complexity based on its distance to a random one of the same size according to a chosen global network complexity metric (see Methods and Fig. 2A). The distribution of its values was also appropriately long-tailed (Fig.

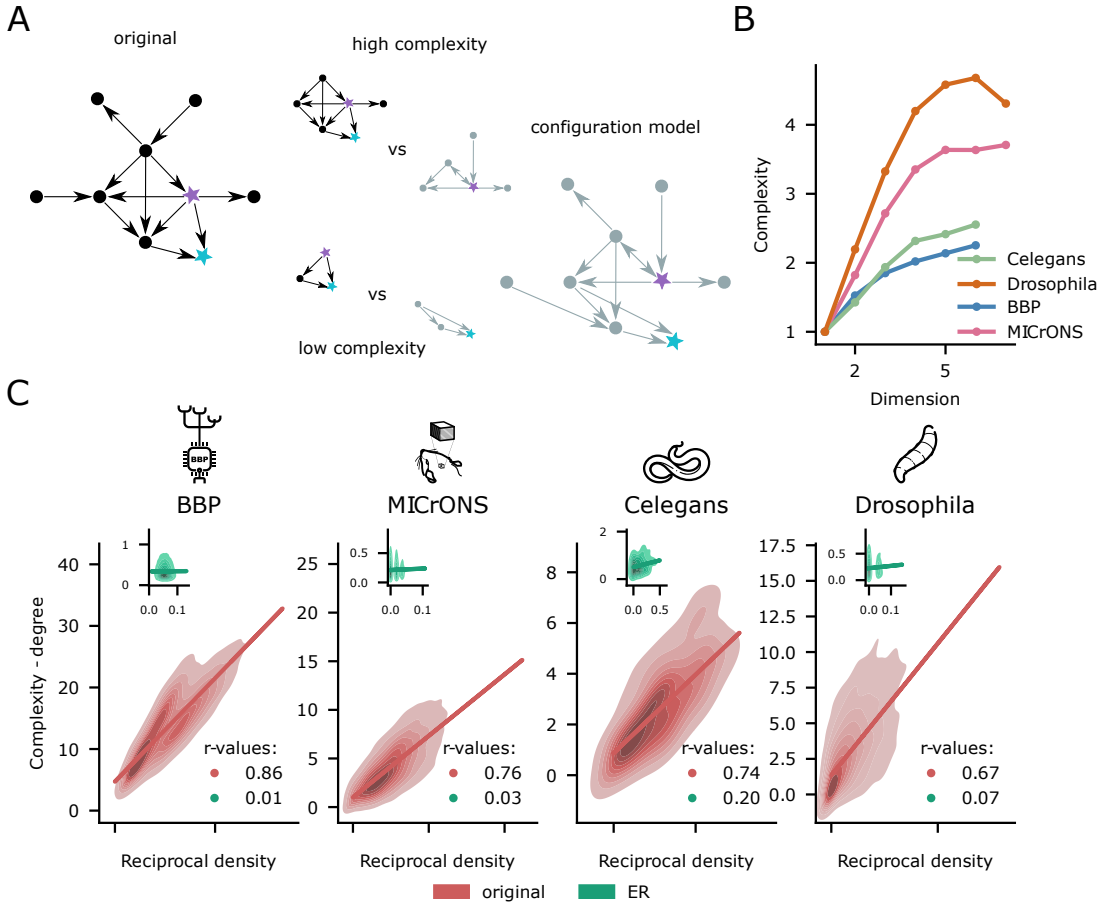

Figure 2: Neighborhood complexity and directed motifs. A: Toy example depicting the concept of neighborhood complexity. The neighborhoods of the purple/blue nodes have high/low complexity respectively. **B:** Average of neighborhood (degree) complexity on nodes that are on simplices of a given dimension (see Methods). **C:** In BNNs reciprocal density is strongly correlated with complexity. Inset, this is not the case for ER graphs.

S1, bottom rows). We computed the average neighborhood complexity across maximal simplices and found that neighborhood complexity increases with dimension for all connectomes (Fig. 2B); indicating that the simplicial structure and neighborhood complexity are strongly related. Given the interaction between the simplicial and reciprocal structures, we hypothesized that a good approximation of the complexity of a neighborhood is its *reciprocal density* i.e., the number of reciprocal connections over the number of neurons. We showed that for BNNs there is a strong correlation between reciprocal density and neighborhood complexity (Fig. 2C) and this relationship is specific to BNNs and not a general property of graphs. Indeed, as can be seen for the absence of a correlation for ER controls (Fig. 2C, Inset) and a weak correlation for Distance controls for the MICrONS and BBP conectomes (not shown). Finally, our results are robust with respect to the choice of network complexity (see Fig. S2).

In summary, a structurally highly diverse set of local subnetworks (i.e., neighborhoods) together form the non-random global structure of BNNs. Moreover, the principle of co-occurrence of reciprocal and simplicial structures at the global level also holds at the local level.

2.3 High complexity neighborhoods promote reliability

We have shown that the architecture of BNNs allows a diversity of neighborhoods and our initially stated hypothesis predicts this should affect function. More precisely, we hypothesized that one of the functions of the highly complex neighborhoods is to promote *reliability*. A neuron is reliable if its firing pattern is consistent (i.e., highly correlated to itself) across several equivalent repetitions (Fig. 3A, S4). We studied this hypothesis for the MICrONS and BBP connectomes, which have co-registered structural and functional data. For MICrONS, calcium traces of a sparse, evenly distributed subset of neurons (up to 2.5% per scan and up to 20% when pooled across scans; see Methods) were reported for 5 classes of naturalistic stimuli in MICrONS[22] (Fig. S3A). Their reliabilities were provided by the authors, for 6 fixed stimuli that were all presented 10 times, as the *Oracle score* of the deconvolved calcium traces over repetitions (see Methods). For BBP, the spike trains of all neurons were reported for eight repeated stimuli (Fig. S3B). Reliability values were computed as the *Gaussian kernel reliability* of the spike trains over repetitions (see Methods and Fig. 3A for two exemplary neurons of high and low reliability).

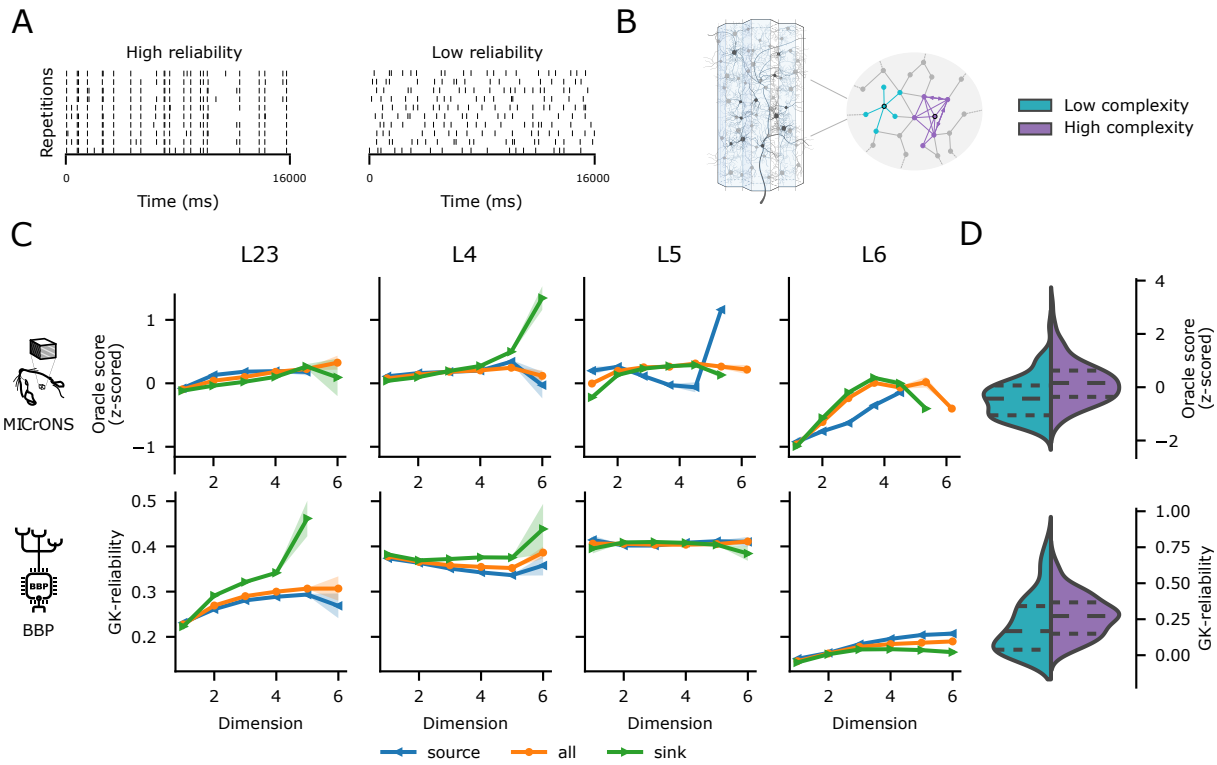

Figure 3: Reliability is shaped by the simplicial structure and network complexity. A: Examples of neurons with low/high reliability in the BBP data. **B:** Cartoon of low (cyan)/high (purple) complexity neighborhoods within a connectome. **C:** Reliability across simplices split by layer for both the MICrONS (Oracle score, top) and BBP (GK reliability, bottom) connectomes. The MICrONS reliability values have been z-scored in order to pool all sessions together maximizing the density of the activity recordings. The roles of the source and target neurons are flipped from the information receiving layers (2-3-4) to the readout layers (5-6). **D:** The centers of the 100 most complex neighborhoods have a mean reliability that is significantly lower than the centers of the 100 least complex ones; with p-values of 3.68×10^{-7} and 1.56×10^{-3} on the Kruskal–Wallis test for MICrONS and BBP respectively.

For a fine-grained perspective, we considered reliabilities separately for neurons in individual layers. We found that reliability increased with cortical depth (moderately for the BBP data, slightly for MICrONS) until layer 5, after which it dramatically dropped in layer 6. Furthermore,

for both datasets the reliability of a neuron increased with the maximal simplex dimension it participates in (Fig. 3C, orange curves), although this increase was weaker when the baseline reliability in a layer was already high, indicating that reliability cannot be raised over a certain ceiling value. Moreover, in layers 5 and 6, typically associated with cortical-cortical and cortical-thalamic outputs respectively, we found that sources of simplices are more reliable than sinks (see Methods, Fig. 3 blue vs green curves); while the opposite is true for the other layers. These trends were clear in the BBP dataset but were inconclusive in MICrONS. This could be related to the increasing sparsity of the calcium imaging data across layers in MICrONS. Indeed, when pooled across sessions, while up to 35% of the neurons are co-registered in layers 2/3, this drops to 15% and 3% in layers 5 and 6. Within a single scan, the data was too sparse to make a layer by layer split; the general trend of reliability increasing with simplex dimension was still observed, however (Fig. S5).

From an even more fine grained perspective, we also contrasted the reliability of individual neurons with their neighborhood complexity for both data sets. Although we found a negligible correlation between neighbourhood complexity and reliability over the entire connectomes (Fig. S6), we found that between the *champions* i.e., those 100 neurons whose neighborhoods are either the most or least complex, there is a statistically significant difference in their reliability (Fig. 3D).

Our observations demonstrated a relationship between neighborhood complexity and reliability. However, these were weak on the individual neuron level. This suggests that the network complexity regulates the reliability of the network as a whole. In order to test this hypothesis, we need to manipulate the overall connectivity and measure the resulting network's activity. This can naturally only be done *in silico* and thus we test this only for the BBP microcircuit.

We employed the recently developed `Connectome-Manipulator` Python framework[44], which allows rapid connectome manipulations of network models in the SONATA format[45] (see Methods). We performed manipulations that both increase or decrease the network's complexity in different magnitudes while keeping the overall density constant (Fig. S8) and we use overall simplex counts as a metric for complexity (Fig. 4A).

We performed four different types of manipulations that *rewire* the model connectome to change its complexity in ways that are consistent with the structural metrics we have shown to shape BNNs across species. That is, they increase or decrease reciprocity or simplex counts (see Methods and Table S1); different manipulations rewire a different amount of connections. Note however, that removing network complexity is always easier than generating it. Indeed, generating networks with the complexity observed is BNNs is an open problem in combinatorial and stochastic topology[32,33].

We found that reducing network complexity always lowered the mean reliability of the network while the opposite happens for increased complexity (Fig. 4A vs B). At the same time, other spiking statistics, such as firing rate distributions, remained unchanged (Fig. 4C). In most cases the change of mean reliability was statistically significant (Fig. 4D1, KW-test for $n = 26567$), and the p-value as well as the size of the change was strongly determined by the percentage of the edges that had been rewired (Fig. reffig:manipulationsD1-inset).

Note that even for the most severe reduction of complexity a small number of highly reliable neurons in a long tail of the distribution remained. Therefore, we tested whether the increase in mean reliability was caused by a global shift or rather by increasing the presence of small populations of highly reliable neurons. This can be captured by the skewness of the reliability distributions, which intuitively is given by the normalized difference between the mean and the median (see Methods). A unimodal distribution has a skewness of 0 if it is symmetric and above/below 0 if it is skewed to the right/left. We found that increasing network complexity

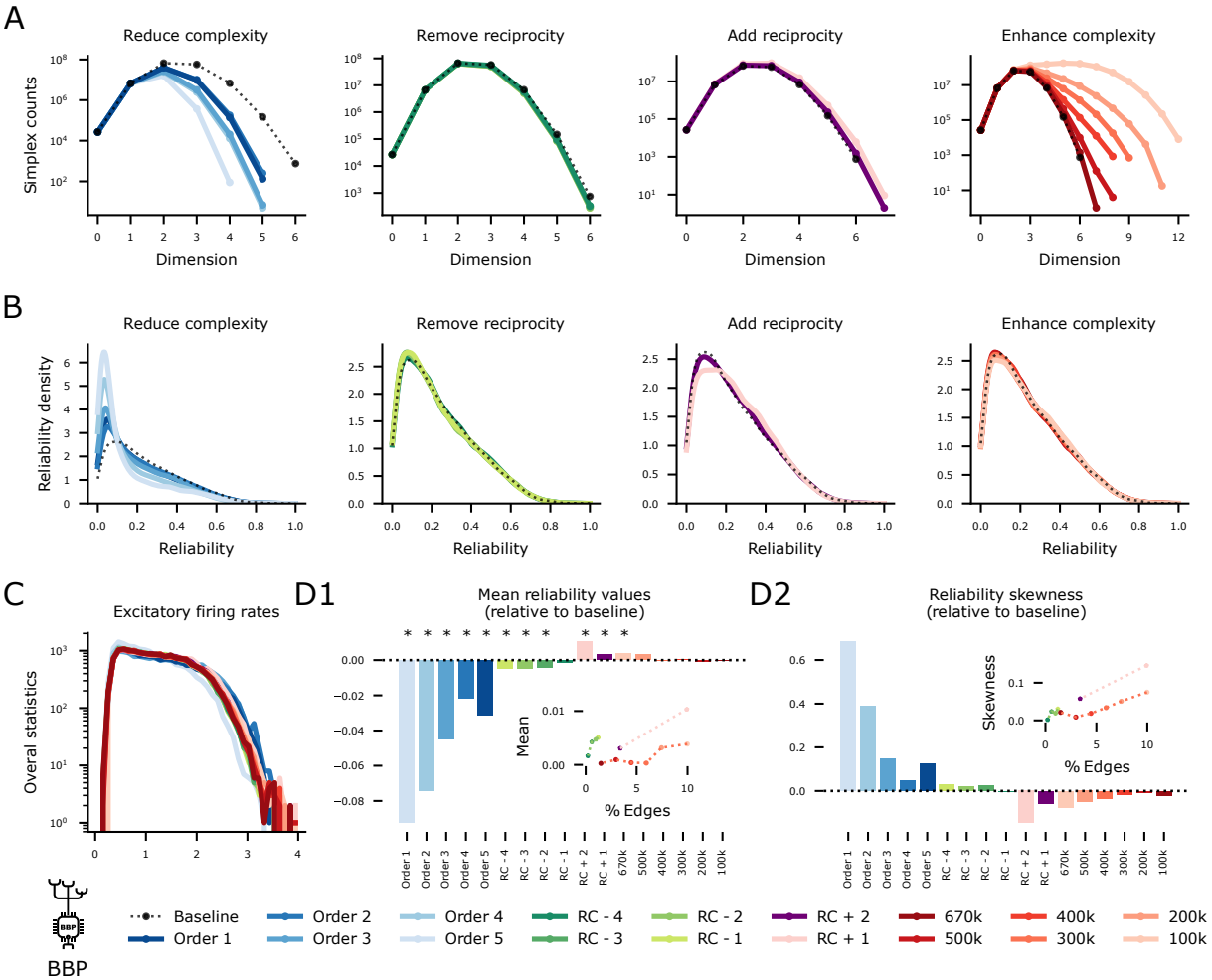

Figure 4: Network complexity enhances network reliability globally. A: Simplex counts of manipulated connectomes. Across all panels, black denotes the baseline values, blue and green reduce complexity and red and purple increase complexity; lighter tones denote stronger manipulations than darker tones. **B:** Distribution of the reliability values of all the neurons in the different manipulated connectomes relative to the baseline. **C:** Firing rates remain stable across all manipulations. **D1:** Mean reliability values of the manipulated connectomes relative to the baseline. Positive values denote an increase of reliability in the manipulated connectome, the $*$ marks manipulations for which the change on the mean reliability is statistically significant (with p-values smaller than 0.009 with $n = 26567$ on the Kruskal–Wallis test). Inset, absolute value of the change of the mean versus the % of the edges that have been manipulated. The reduced complexity models are not shown in the inset since virtually 100% of the edges are modified in these manipulations. **D2:** As D1 but for skewness of the distribution rather than the mean value.

decreases the skewness of the reliability distribution (Fig. 4D2), showing that the mean reliability increase is due to a global network effect.

In summary, network complexity increases reliability locally in two ways: On the node level, reliability increases with simplex dimension; on the subnetwork level, the most complex neighborhoods have significantly more reliable centers than their least complex counterparts. However, the strongest effect of network complexity is to regulate the reliability of the network as a whole.

2.4 Low complexity neighborhoods are efficient

So far, we have shown that BNNs have a wide range of neighborhood complexities and that the the most complex neighborhoods promote reliability; they have also been shown to promote robustness[6,7]. This leads to the question of what is the purpose of the least complex neighborhoods? Our initial hypothesis states that they encode information more efficiently, in spike trains that are less mutually correlated, thus providing a diversity of roles for the neurons in the network. A related spectrum of functional roles has been observed experimentally, where neurons can differ in their activity correlation to the overall firing of the population, i.e., their *coupling coefficient*, ranging from strongly coupled *choristers* to weakly coupled *soloists*[46,47].

Indeed, we found for both MICrONS and BBP that the simplicial structure shapes the value of the coupling coefficient, where neurons that belong to higher dimensional maximal simplices have higher coupling coefficients (Fig. 5A). This shows that soloists/choristers are in the lower/higher simplicial dimensions of the connectomes. Most neurons belong to a high dimensional maximal simplex, $\sim 95\%$ and $\sim 98\%$ of all neurons belong to simplices of dimension 5 and higher for MICrONS and BBP respectively. This provides an explanation for the finding of Okun et al.[46] that choristers outnumber soloists. Moreover, they also found that the coupling coefficient increases with the amount of synaptic input. By showing that sinks of simplices are more coupled than sources, we confirm this finding and extend it by showing that the effect is even stronger if the presynaptic population is highly recurrent.

The concept of soloist and choristers addresses the coupling of activity on the level of single neurons. In order to explore the concept of efficiency as a metric of relative information capacity, we study the level of redundancy in the activity of neighborhoods. A highly efficient neighborhood must have mostly uncorrelated activity, thus maximizing the amount of information it can encode in its activity pattern. Hence, we define the *efficiency* of a neighborhood as the fraction of its neurons required to explain 90% of the variance of its correlation matrix[48–53] (see Methods, Fig. 5B).

The sparsity of the MICrONS functional data set makes it impossible to estimate this metric for that connectome. Thus, we restrict our analysis to the BBP data set. We found that efficiency decreases with the dimension of maximal simplices (Fig. 5C-left) and generically sinks of neurons are more efficient than sources. From a more fine grained perspective, we found that across layers neighborhoods centered on sink neurons were also more efficient than neighborhoods of source neurons and that globally neighborhoods with centers in layer 6 were more efficient than other layers (Fig. S7C). The later mirrors an opposite trend we observed for reliability. On the other hand, as expected, highly efficient neighborhoods tend to have soloists at their center and it is anti-correlated with reliability and neighborhood complexity (Fig. 5C-right).

The observed trends should make low complexity neighborhoods also efficient readout populations for information processed in a local circuit. We tested this in a simple stimulus classification task. For this, we simulated evoked activity in response to a random stream of the 8 stimuli used before for characterizing reliability (Fig. S3B). Each stimulus was presented 100 times in random order with an inter-stimulus interval of 200 ms. We then tried to determine the stimulus present in a given time window from the spiking activity of a subpopulation of neurons. In order to test the effect of neighborhood complexity, we selected the subpopulations as neighborhoods at different points along the complexity spectrum.

Specifically, we used the neighborhood-based classification pipeline developed in Reimann et al.[54] and Conceição et al.[55]. Briefly, it selects 50 neighborhoods either: at random, or by the ones that maximize or minimize a given network metric from Table S1. We call their original pipeline the *single selection* procedure. This will sample neighborhoods that are either

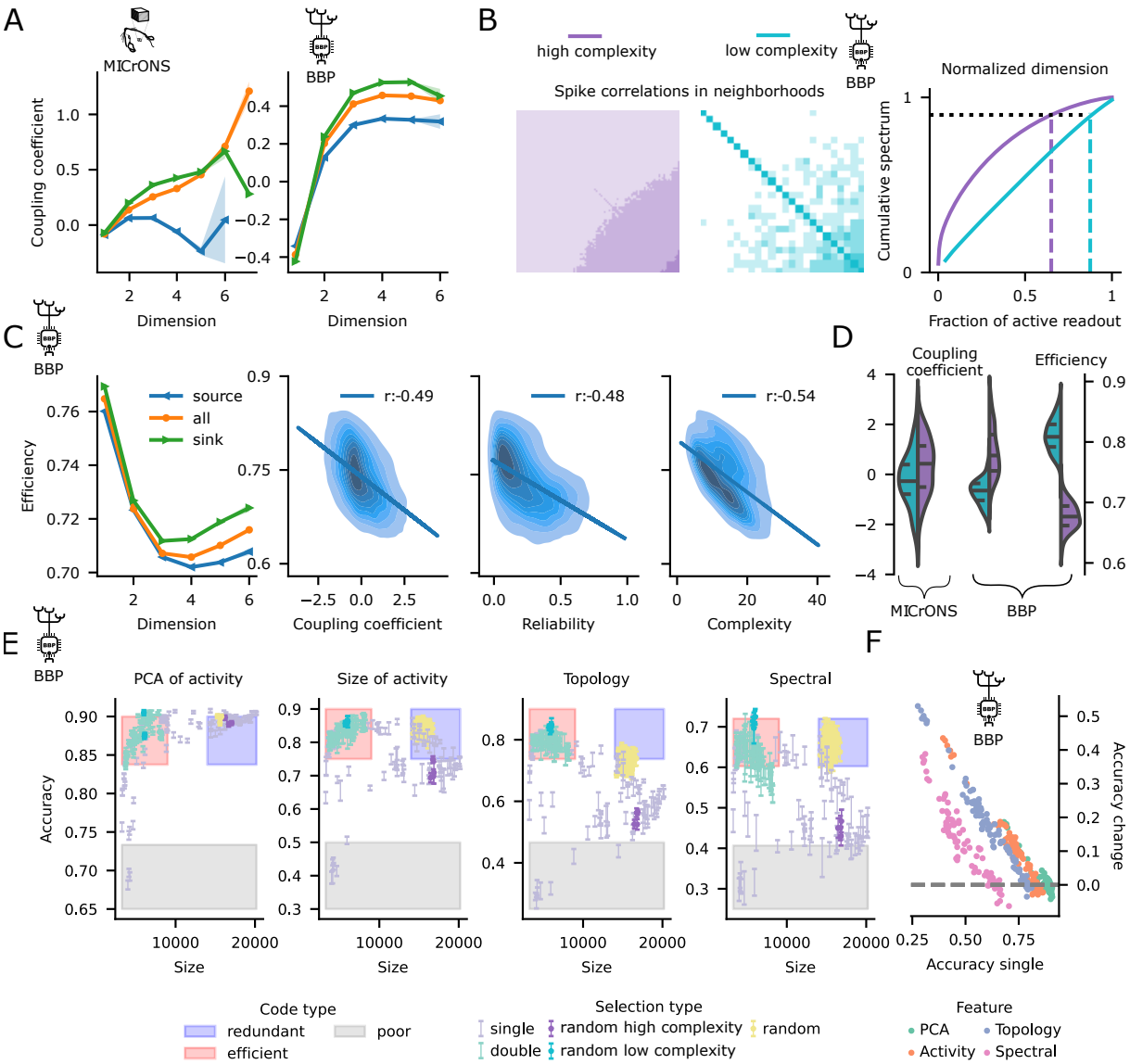

Figure 5: Low complexity neighborhoods are more efficient. A: Coupling coefficient (z-scored) increases with maximal simplex dimension for MICrONS (left) and BBP (right). Sinks of neurons (green curves) are more coupled than sources (blue) and the overall population (orange). **B:** Left: correlation matrices of the spike trains of the lowest/highest complexity neighborhoods of the network in the BBP data. Right: their corresponding normalized spectra from which their *efficiency* is determined. **C:** Efficiency for the BBP data, from left to right: Efficiency decreases with maximal simplex dimension colors as in A. Efficiency is anti-correlated to the coupling coefficient, reliability of the central neuron and neighborhood complexity. **D:** Left: The centers of the 100 least complex neighborhoods have a mean coupling coefficient that is significantly lower than the centers of the 100 most complex ones; with p-values of 1.66×10^{-3} and 8.25×10^{-28} on the Kruskal–Wallis test for MICrONS and BBP respectively. Right: The opposite effect is observed for Efficiency, with a p-value of 2.60×10^{-34} (BBP data only). **E:** Accuracy of the classification task versus size of the subpopulation used for four different featurization methods. Less complex neighborhoods (light blue) consistently give provide higher accuracy with smaller subpopulations; even when these are chosen at random (cyan). **F:** The change in accuracy in the double selection process (see Methods) is anti-correlated to the accuracy of single selection; showing that low complexity is a sufficient condition for efficient coding.

a generic representative of a circuit neighborhood, or structural outliers in one way or another. We explicitly consider outliers, as the values of network metrics over neighborhoods are long-tailed[38,37] (Fig. S1) and random sampling may miss points from the extreme ends. Then, the spiking activity of the union of neighborhoods is reduced to a set of time series that we call the *feature vectors*. This featurization is done in two ways: either by applying principal component analysis to the binned spike trains (*PCA method*; bin size 20 ms); or by calculating the value of one of the network metrics listed in Table S1 of the *active subnetwork* of each neighborhood and in each time step (see Methods). That is, we reduce the dimensionality of the spike trains of the subpopulation in different ways to ensure the robustness of our results. Finally, a simple linear classifier is trained to identify the stimulus identity from these feature vectors.

Unsurprisingly, we found that the classification accuracy tended to be higher when the size of the union of the neighborhoods was larger (Fig. 5E, purple bars and Fig. S10). However, we were primarily interested in efficient coding, not its absolute quality. We considered a neighborhood to be more efficient than another if it reached the same or higher classification accuracy with fewer neurons. As we hypothesized that less complex neighborhoods are more efficient, we executed a *double selection* procedure, where we first selected the 1% least complex neighborhoods and ran the full pipeline described above on those neighborhoods only. When considering the size of the neighborhood union against classification accuracy, we found the following: First, double selection resulted in smaller neighborhoods. Second the data points associated with double selection had comparable or better classification accuracy despite the smaller neighborhoods (Fig. 5E, light blue and cyan bars, red-shaded region). They were therefore more efficient according to our definition.

The overall level of classification accuracy depended on the network metric used to calculate the feature vectors (Fig. S10). For a minority of metrics (9 out of 32) the classification was poor and did not improve for double selection. In these cases, less than $\sim 4.5\%$ (generally less than $\sim 2\%$) of the entries of the feature vectors where nonzero (Fig. S11. That is, the featurization was degenerate since virtually no input information was given to the classifier. For some selection parameters, classification accuracy was already high in the single selection paradigm. In those cases, it did not further increase for double selection. Instead, double selection ensured high accuracy where accuracy for single selection was low (Fig. 5F). This demonstrates that low complexity is a sufficient condition for efficient coding.

In summary, low complexity neighborhoods encode information most efficiently, and this encoding is made robust by them being interconnected with highly reliable neighborhoods. Furthermore, layers associated with cortical outputs increase efficiency at the cost of reliability along their simplicial structure (i.e., from source to sink).

3 Discussion

We have further extended our knowledge about the immense diversity of structure found in neuronal networks of different species and at cellular resolution. While the presence of wide, long-tailed degree distributions were known, we found similarly long-tailed distributions for a variety of network metrics, in particular those capturing participation in the higher-order structure of the network. This persisted even in the face of normalization with strong control models. Further, we found that participation of neurons in the higher-order structure was correlated with previously characterized overexpression of bidirectional connectivity[17,46]. This gives rise to subnetworks centered on individual neurons (i.e., neighborhoods) that differ greatly along a spectrum of topological complexity. We linked this spectrum of complexity to circuit function, demonstrating that complex subnetworks provide a robust code and play a supporting role, globally increasing the reliability of the neuron responses in evoked activity. Conversely, less complex

subnetworks encode information with little redundancy and hence, efficiently. These results remained consistent, even when the measure of topological complexity was changed (Fig. S2). Going beyond abstract theory, we confirmed the efficiency of the subpopulations in a simulated stimulus classification experiment.

The contrast between reliability-enhancing and efficiently-coding neurons is related to the previously characterized contrast between "soloists" and "choristers"[47,46]. We indeed confirm their results between connectivity and population coupling, and extended it to demonstrate the importance of higher order structures. The importance of local connectivity for promoting reliable responses has also been found before[6], but we were able to characterize which connectivity motifs relate to it. Similarly, functional cell assemblies, which provide robust coding through redundancy, have been demonstrated to emerge more efficiently, that is, through fewer synaptic connections, in networks with complex higher-order structure[27]. In summary, we confirm, extend and link several known structural and functional features of biological neuronal networks into a single, coherent and quantitative theory. The crucial aspect linking these aspects is consideration of the higher-order structure of the network, i.e., going beyond pairwise interactions.

Many of our conclusions were drawn from the analysis of electron-microscopic connectomes, primarily the MICrONS data[22]. While reconstructed with boundary-pushing methods, it is still imperfect. To link a synapse to its pre-and post-synaptic neurons, the associated dendritic and axonal volumes must be correctly segmented along the entire path to the soma. Errors in this process are more likely in the thinner axon and would hence result in a loss of efferent connections or assignment of the wrong presynaptic neuron. It is virtually impossible that the higher-order structure we characterized and linked to function is a result of such random perturbations, as it is already difficult to build intentionally[32,33]. Our conclusions relating to efficiency of coding are based on simulations of a morphologically detailed model of cortical connectivity[23]. They could be confirmed in the future when more or more densely sampled electron-microscopic connectomes with co-registered activity data become available. In the meantime, the strong qualitative agreement between MICrONS and BBP data in virtually all characterized trends indicates the validity of our model predictions. Additionally, going beyond correlation and demonstrating causation between higher-order structure and neuron function requires use of a detailed model, as the very fine-grained connectivity manipulations required are not feasible experimentally. They are also difficult to do in more simplified network models, as they typically lack the characterized higher-order structure in the first place, and it is an unsolved problem to generate such structure without using neuron morphologies[32,33]. This demonstrates the utility of pairing experimental data analysis with carefully considered simulation experiments, such as this study.

To show that the higher-order structure and neighborhood complexity are causative and not merely correlated with reliability and efficiency, we ran additional simulations where we increased or decreased the amount of network complexity and studied the global effect on reliability; the need to do this has been pointed out before in Reid et al.[56]. We found that manipulations changed the mean reliability significantly and in the expected directions. Additionally, for a given rewiring algorithm, the amount of change is dependent on the fraction of connections rewired, as expected. The strength of the effect was stronger for algorithms that decreased complexity, confirming the idea that it is easier to destroy than to build the intricate wiring characterizing cortical circuits. Additionally, the manipulations changed the shape of the distribution of reliability values, such that increasing / decreasing mean values were paired with decreasing / increasing skewness. This shows that the network structures we characterized increased the overall reliability of the entire population rather than only that of specific neurons.

On the other hand, there were only a small number of high efficiency neighborhoods providing high classification performance.

Beyond the general ideas outlined above, we also found more specific results relating to the roles of cortical layers that align with their known roles as inputs or output. Our notion of higher-order interactions is based on the presence of simplices, which are motifs with strong directionality, i.e., they have an input or *source* side and an output or *sink* side. In the simulation data we found that, for neurons in layers 2, 3 and 4 efficiency decreased and reliability increased from the source to the sink of a simplex, while in layer 6 this was reversed. Neurons in layer 5 behaved as layer 6 with respect to efficiency and the picture was non-conclusive for reliability, although they were overall the most reliable population. The *in vivo* data was more sparsely sampled than the simulated data, leading to noisier results. In fact, the data were too sparse to either estimate efficiency or to analyze reliability separately for individual layers within a given recording session. However, the per-layer analysis of reliability could be performed when pooling data across sessions but the results where not conclusive as the number of sample neurons decreased towards the deeper layers. Layer 5 is associated with extensive intratelencephalic, i.e., cortico-cortical outputs and layer 6 with cortico-thalamic outputs[4,5]. Therefore, our interpretation is as follows: After inputs into a circuit arrive canonically in layer 4, local connectivity structure is optimized to enhance reliability while progressing activity forward. Within output layers, simplicial structure still increases reliability and enhances efficiency specifically of the output neuron population.

This view is compatible with the theory of "reservoir computing" or "liquid state machines"[57], which, in the context of neural circuits, posits that circuit inputs are first projected into a high-dimensional activity state and from there projected to a lower-dimensional readout. Seen in this context, we characterized connectivity mechanisms that increase the robustness of the higher-dimensional projection (ironically reducing the dimensionality of that state), and increase the amount of information contained in the readout. In summary, our hypothesis predicts increasing robustness as the goal for the internal representation of a local circuit, and increasing efficiency as the goal for its output, with specific structures of connectivity associated with both goals and embedded into a single network that promotes the reliability of the whole system. Our work can therefore be used in the future to link also this theory to the topics of non-random connectivity, efficiency and soloist/chorister spiking dynamics.

Our view sees trial to trial variability as a problem that needs to be overcome for reliable cortical function. This appears to be in contrast to existing theories where variability explicitly encodes uncertainty about a stimulus[58,59]. However, we believe the views are compatible. First, the theoretical work sees uncertainty encoded primarily on the population level[59], while we measured reliability on the single cell level. Second, we only investigated the prevalence of variability, not its structure and how it relates to stimuli. Finally, we found remaining uncertainty in the responses of even the most reliable cells. Furthermore, our results demonstrate a wide spectrum of reliability values, and that the mechanisms to reduce it are not uniformly deployed. It is therefore possible that they are deployed in ways that implement the theories of explicit encoding of uncertainty. Investigating this would constitute interesting follow-up work.

Experimental research has demonstrated several ways in which neuronal connectivity is non-random with significant higher-order structure[18,21,28,29,31]. Moreover, small models with simple dynamics have shown the relevance of higher-order motifs in the observed activity[60–62] while *in silico* research has showed that the complexity is largely driven by neuronal morphologies[63,64,20].

At the same time, in the critical field of computational neuroscience, the bulk of research is conducted in models with simple connectivity based only on pairwise statistics that do not capture the entire complexity of the structure. This is in part because due to the difficulty of building such complex networks; but also due to the need for a framework to quantitatively and systematically describe the higher-order structure that allows to demonstrate how this structure matters for the function of a neural circuit, if at all. Our work addresses this oversight, providing a theoretical framework, based on counting simplex motifs, that links and unifies the diverse aspects of non-random higher-order structure that have been found before. In contrast to global metrics of network complexity, e.g., small-world coefficient or assortativity, simplex counts give a finer grained metric further allowing to give notions of node and edge centrality connected to them which can additionally be used to study biological predictors for functional properties. Moreover, there are several advantages of counting these motifs over counting all possible motifs on n nodes, e.g., counting triads for $n = 3$. First, possibly due to their feedforward nature, simplices have been shown to be functionally meaningful[19,27,65]. Second, counting n-simplices can be done systematically for all dimensions, and the maximum dimension attained (at a given graph density) is an indicator of network complexity. Third, counting n-simplices can be efficiently implemented as opposed to counting the "core motifs" or "robust motifs" of Parmelee et al.[61] and Curto et al.[62], for which even listing them concretely above dimension 5/6 remains an open problem. Moreover, whilst counting n-simplices and counting directed cycles on n nodes have the same worst-case complexity, increasing exponentially with the number of nodes[66]; in practice, counting n-simplices can be computed more efficiently, particularly in large sparse graphs.

Taken together, this and the cited previous research demonstrate that the structure-function relation at cellular resolution can only be understood in the context of the biological constraints on both. Structurally, the immensely dense packing of neuronal wiring shows that a given function will be implemented with the least number of synapses possible[67]. Functionally, due to the physical movement of neurotransmitters at synapses that need to be restored a synapse cannot always respond perfectly reliably (or only at great expense in terms of space and metabolic cost), leading to variability and noise that needs to be overcome. One way of doing this is adding redundant connections to the graph structure of the connectome, but which is at odds with efficient wiring and coding. We posit that biological neural networks solve this apparent contradiction through a wiring that promotes network diversity. If these constraints are not taken into account, we cannot understand the implementation of the neural code in synaptic connectivity.

4 Methods

4.1 Structural data sets

We analyzed the connectivity in the following four openly available data sets recorded at cellular resolution. We used the `ConnectomeUtilities` Python package[68] to turn these data sets into the common "ConnectivityMatrix" representation format for further analysis.

C. elegans[34] Electron microscopy reconstruction of the full brain connectivity of the adult C. elegans (Dataset 8; last stage). The data set consists of 324 neurons and 2186 connections formed by 7970 chemical synapses. We did not include electrical synapses in our analyses. Additionally, we discarded 105 isolated neurons of the connectome, i.e., neurons without incoming or outgoing connections.

Drosophila larva[35] Electron microscopy reconstruction of the full brain connectivity of the Drosophila larva. We included both hemispheres but only considered "axo-dendritic" con-

nections. The resulting data set consists of 2956 neurons and 63,545 connections formed by 234,958 synapses.

MICrONS[22] Electron microscopy reconstruction of a cubic millimeter volume spanning all 6 layers of mouse primary visual cortex and three higher visual areas. A reformatted version for easy use with `ConnectomeUtilities` is available on Zenodo[69]. We excluded the boundaries and restricted the data set to excitatory types. The resulting connectome consists of 9619 neurons and 520,477 connections formed by 610,261 synapses.

BBP[23] Morphologically detailed model of the rat somatosensory cortex[23], consisting of more than 210K biophysically detailed neurons belonging to 55 different morphological and 207 different morpho-electrical types. The neurons are connected by over 400M synapses with probabilistic transmitter release and six distinct forms of short-term dynamics of depression and facilitation. The whole circuit can be divided into seven hexagonal columns. To avoid any edge effects, we restricted all analyses (connectivity and activity) to neurons within the central hexagonal column. This data set is available on Zenodo[70] and can be loaded with `ConnectomeUtilities`. Within that column, we only considered the sub-circuit of 6,717,001 connections between 26,567 excitatory neurons. In the context of connectome manipulations (see Section 4.8), we refer to this data set as *baseline connectome*.

4.2 Activity data set for MICrONS

We analyzed functional data from the MICrONS dataset[22]. The open source dataset contains calcium traces and deconvolved spike trains of a sparse set of neurons, which are co-registered to the structural dataset. We analyzed data from eight recording sessions from the v661 version of the "minnie65_public" release. The sessions were selected based on two criteria: at least 1000 neurons were scanned and at least 85% of them were co-registered to the structural data. Neurons with a non-unique identifier (2.8% neurons/session on average) were filtered out. Averaged across sessions, 2% of the neurons in the connectome (243 neurons/session) had activity data co-registered to them. Moreover, as calcium imaging is biased towards the superficial layers, the layer-wise distribution of co-registered cells is not uniform, but as follows: layer 23: 36.5%, layer 4: 44.9%, layer 5: 15.3%, and layer 6: 3.3%. Spike trains of neurons restricted to the structural connectome outlined above (243 neurons/session on average, 1817 unique neurons in total, 2% and 15% of all cells respectively) were used to calculate coupling coefficients (see below). As the dataset already contained oracle scores (see below) we did not calculate reliability from the raw spike trains.

4.3 Activity data set for BBP

We ran circuit simulations of the morphologically detailed model of the rat somatosensory cortex[23], which was converted to the SONATA format[45] in order to create manipulated versions of its connectome (see Section 4.8). We simulated the entire circuit (original and manipulated versions) using the CoreNEURON simulator[71] and recorded the spiking activity of the network in response to a series of stimuli applied to the thalamocortical input fibers. As in Reimann et al.[54], the 2170 input fibers were partitioned into 100 clusters of adjacent fibers. Ten such clusters were randomly drawn to form a spatial stimulus pattern. For each cluster belonging to such a pattern, a spike train with a firing rate of 75 Hz was independently generated by an adapting Markov process (with 100 ms adaptation time constant) for 75 ms, followed by a 125 ms blank period with 0.5 Hz firing. We used a random sequence of eight such spatial patterns in two simulation protocols:

Reliability protocol We applied a total number of 80 stimuli (i.e., 10 repetitions per pattern) during 16 s. We reran the experiment 10 times with different simulator seeds using the exact same (random) sequence of stimulus patterns and input spike trains.

Classification protocol We applied a total number of 800 stimuli (i.e., 100 repetitions per pattern) during 160 s. We used a single simulator seed without repeating the experiment.

4.4 Random networks

This section contains definitions of the different random connectivity control models and how these were generated. A summary of these is listed in Table 1.

Table 1: Types of random graphs. The lower a model is in the table the closer it is to the original graph.

Label	Model name	Property maintained
ER	Erdős–Rényi	Overall density of connections
CM	Directed configuration model	Sequence of in-/out- degrees
Dist	Distance dependent	Density of connection at a given distance
Und	Underlying	Underlying undirected graph
rcSh	rc-Shuffled	Subgraph of non-reciprocal connections

The first three controls of Table 1 modify the global structure of the network whilst keeping one property fixed. They are listed in order of complexity. The *Erdős–Rényi* controls (ER) are random directed networks, where each edge is added independently at random with a fixed probability p given by $p = \frac{E}{N(N-1)}$, where E and N are the number of edges and nodes in the original network, thus keeping the global connection probability fixed.

The *directed configuration model* controls (CM) are random networks with (approximately) fixed in- and out-degree sequences, i.e., the vectors whose entries are the in- and out-degrees of all the nodes in the network. To construct such controls we wrote the original network in coordinate format i.e., we wrote its edges as two vectors *sources* and *targets*, such that each pair $(sources[i], targets[i])$ was a directed edge in the original network. We then shuffled the entries of both vectors independently, which gave a new directed network, with the same degree sequence as the original. Note that this new directed network might contain loops (edges of the type (a, a)) or parallel edges (multiple edges with the same source and target). We call these degenerate edges and we removed them. The effect of this removal is that this construction only approximated the original degree sequence. However, the density of degenerate edges tends to decrease as the number of nodes increases[72], so the approximation was close. Indeed, while on average $\sim 5\%$ of the edges were lost for the CM controls of C. elegans, only $1.08\%, 1.15\%$ and 0.875% of the edges were lost for Drosophila, MICrONS and the BBP data sets respectively.

The *distance dependent model* controls (Dist) are random networks where the edges are added independently at random with a probability that is exponentially decreasing with distance. More precisely, for any pair of neurons, let d be the Euclidean distance between the position of their cell bodies. Then, their probability of connection is given by

$$p(d) = \alpha \cdot e^{-\beta \cdot d},$$

where α (the probability at distance zero) and β (the decay constant) were determined from the connectivity of the original graph. In order to do so, we computed the pairwise distance between all pairs of nodes in the network and binned them into groups according to this distance. The connection probabilities were estimated in each bin by dividing the number of existing edges by

the number of total pairs in that bin. The model coefficients were then determined by fitting the exponential probability function $p(d)$ to these data points. We did this only for the MICrONS and BBP data sets, obtaining $\alpha \approx 0.067, \beta \approx 1.18 \times 10^{-2} \mu m^{-1}$ and $\alpha \approx 0.111, \beta \approx 7.05 \times 10^{-3} \mu m^{-1}$ respectively. For C. elegans and Drosophila the Euclidean distances between cell bodies might not be the most relevant factor determining connection probability.

The last two controls in Table 1 are used to control for the possible location of reciprocal connections, while keeping the number of simplex counts close to those of the original graph (specially for large sparse graphs). The *Underlying* controls (Und) are random networks with the same underlying undirected graph as the original and the same number of reciprocal connections. To generate such controls, we first extracted the underlying undirected graph of the original network i.e., the undirected graph with the same nodes as the original and an undirected edge between any two pair of nodes that are connected by a reciprocal or unidirectional edge. We then chose an ordering of the nodes and oriented all edges of the undirected network to go from a smaller to a larger node. Finally, we added the same number of reciprocal connections as in the original graph at random.

The last control, the *rc-Shuffled* control (rcSh) is the tightest control of all. These are random networks where all of the structure of the original network is kept, except the location of the reciprocal connections. To build such a control, for every reciprocal edge in the original network we chose one direction to delete at random. Then we introduced back at random the same number of reciprocal connections along the existing edges.

4.5 Network metrics

In a network G the *in-degree* and *out-degree* of a node v is the number of incoming and outgoing connections from v, respectively, and the *total degree* of v is the sum of both. We call the *reciprocal degree* of v the number of reciprocal connections on v, i.e., the number of nodes in G that that map to and from v.

An *n-simplex* in G is a set of $n+1$ nodes which are all to all connected in feedforward fashion. That is, there is a numbering of the nodes $0, 1, \ldots n$ such that whenever $i < j$ there is an edge from i to j in the network G. We call n the *dimension* of the simplex, the numbering $0, 1, \ldots n$ the *position* of the nodes in the simplex and the edges

$$0 \to 1 \to 2 \to \ldots \to n-1 \to n$$

the *spine* of the the simplex. A simplex is *maximal* if its nodes and edges are not contained in a higher dimensional simplex.

We weight the edges of G with the maximal dimension amongst all the simplices it participates in and we denote by G_k the subgraph of G on the edges with weight $\geq k$. That is, the edges in G_k belong to at least one simplex of dimension k. In particular $G_1 = G$, since every edge is a 1-simplex. We then compute the *reciprocity of G at dimension k*, denoted rc_k, as the percentage of reciprocal edges of G_k

$$\mathrm{rc}_k = \frac{\text{number of reciprocal edges of } G_k}{\text{number of edges of } G_k} \times 100.$$

We plot the curves $k \mapsto \mathrm{rc}_k$ for all connectomes and their corresponding controls in Figure 1D.

The simplicial structure of G can be used to compute (weighted) averages of any node based property as follows. Let N be the number of nodes of G and $\vec{v} \in \mathbb{R}^N$ be a vector given by a node based property, e.g., node degree or reliability. For any dimension k and any possible position $0 \leq pos \leq k$ within a k-simplex, we denote by $\mathrm{mean}(\vec{v}, k, pos)$ the mean of the values of $\vec{v}$ of

the nodes in position *pos* of k-simplices, counted with multiplicity. More precisely, if $s^1, \ldots, s^t$ is a list of all k-simplices, let $\vec{v}(s^i_{pos})$ be the value of $\vec{v}$ of the node in position *pos* of s^i, then we define $\text{mean}(\vec{v}, k, pos)$ as the mean

$$\text{mean}(\vec{v}, k, pos) = \text{mean}\{\vec{v}(s^1_{pos}), \vec{v}(s^2_{pos}), \ldots, (s^t_{pos})\}.$$

We denote by $\text{mean}(\vec{v}, k, all)$ the weighted mean of $\vec{v}$ on the nodes that are in k-simplices (counted with multiplicity) regardless of position, i.e.,

$$\text{mean}(\vec{v}, k, all) = \text{mean}\{\vec{v}(s^1_{pos}), \vec{v}(s^2_{pos}), \ldots, (s^t_{pos}) | 0 \leq pos \leq k\}.$$

We plot the curves $k \mapsto \text{mean}(\vec{v}, k, all)$ in Figures 2B, 3C, 5A/C for different structural and activity properties of the nodes. In the last three cases, we further plot $k \mapsto \text{mean}(\vec{v}, k, pos)$ for $pos = 0$ or $pos = k$ i.e., the source respectively target neurons of a simplex.

4.6 Neighborhood complexity

Let G be a network and v be a node in G. We call the *neighborhood of* v the subgraph of G on all the nodes in G that are connected to v and the edges between them. We denote this subgraph by N_v and we call v its *center*. The *size of* N_v is the number of nodes in it. In the absence of reciprocal connections, the size of N_v is the total degree of v plus 1. However, in general we have that

$$\text{size}(N_v) = \text{total degree}(v) + 1 - \text{reciprocal degree}(v).$$

The complexity of N_v is determined by how similar or different it is from a random subgraph of a similar size. In order to quantify this, we generate a configuration model control of G. We denote this control $\hat{G}$ and by $\hat{N}_v$ the neighborhood of v in $\hat{G}$. Note that by construction N_v and $\hat{N}_v$ are of similar size, particularly for large sparse graphs. Thus, we define the *complexity of* N_v as the distance between N_v and $\hat{N}_v$ according to some global network metric (see Fig. 2A, where G and $\hat{G}$ are in black and grey respectively). First, we consider a metric that takes into account only pairwise interactions and define a degree based complexity metric as

$$\text{complexity - degree}(N_v) = \text{W}(D_{N_v}, D_{\hat{N}_v}),$$

where D_{N_v} and $D_{\hat{N}_v}$ are the total degree distributions of N_v and $\hat{N}_v$ respectively and W is the Wasserstein distance between them, a classical distance between probability distributions[73]. Second, we consider a metric that also takes into account the higher order structure and define a simplicial complexity metric as

$$\text{complexity - simplicial}(N_v) = \text{d}(\overline{SC_{N_v}}, \overline{SC_{\hat{N}_v}}),$$

where $\overline{SC_{N_v}}, \overline{SC_{\hat{N}_v}}$ denote the simplex counts of N_v and $\hat{N}_v$ respectively linearly normalized to jointly lie between 0 and 1, and d denotes Euclidean distance. These two metrics are strongly correlated to each other (Fig. S2D).

4.7 Reliability

We measured the reliability of spike trains to repeated presentations of natural movies in the MICrONS dataset, with the *Oracle score* reported by MICrONS[22]. The score was computed as the jackknife mean of correlations between the leave-one-out mean across repeated stimuli with the remaining trial. The sparsity of the activity data set required us to pool Oracle scores across sessions for some analyses. Since the baseline Oracle scores have large variability across

sessions (Fig. S5A), we first z-scored the Oracle scores in each session and then average them across sessions.

To measure the spike timing reliability of a neuron in the BBP data set, we used the notion of *Gaussian kernel reliability*[74] across repetitions of the same experiment. The repetitions were obtained from repeated simulations with different simulator seeds. Specifically, for a set of spike trains of a given neuron obtained from repetitions of the same experiment, we first convolved the spike trains with a Gaussian kernel with $\sigma = 10\,\text{ms}$. Since reliability is known to be confounded by the firing rate of a neuron[75], we removed this dependency by *mean-centering* the data (see Fig. S4), i.e., by subtracting the mean of the convolved signal. Then, for each neuron we computed the cosine similarity between all pairs of repetitions of mean-centered signals and averaged the resulting values.

4.8 Connectome manipulations

We employed the recently developed `Connectome-Manipulator` Python framework[44], which allows rapid connectome manipulations of network models in the SONATA format[45]. The framework supports the creation of new connectomes (referred to as wiring), transplantation of connectivity features from other connectomes, and manipulations of existing connectomes (referred to as rewiring) on the level of connections and synapses. It also implements a set of stochastic (simplified) connectivity models based on neuron classes and locations in space whose parameters can be fit against existing connectomes, allowing to replace an existing connectome by equivalent connectomes at different levels of complexity.

There were four categories of manipulations done on the baseline connectome: *remove reciprocity, increase reciprocity, reduce complexity, enhance complexity.* These are listed in Table 2 and we describe them explicitly here. All manipulations were restricted to connections between excitatory neurons in the central hexagonal column of the circuit.

Table 2: Types of connectome manipulations

Label	Modification category	Description
$RC-1$	Remove reciprocity	Remove reciprocal connections in maximal dimension $n \geq 5$
$RC-2$	Remove reciprocity	Remove reciprocal connections in maximal dimension $n \geq 5$ and 50% at $n = 4$
$RC-3$	Remove reciprocity	Remove all reciprocal connections in dimension $n \geq 4$
$RC-4$	Remove reciprocity	Remove all reciprocal connections
Order 1	Reduce complexity	Uniform probability of connection
Order 2	Reduce complexity	Distance-dependent probability of connection
Order 3	Reduce complexity	Bipolar distance-dependent probability of connection
Order 4	Reduce complexity	Offset-dependent probability of connection
Order 5	Reduce complexity	Position- and offset-dependent probability of connection
$RC+1$	Add reciprocity	Reciprocal connections $\times 2$ across simplices
$RC+2$	Add reciprocity	Reciprocal connections $\times 4$ across simplices
$100K$	Enhance complexity	Rewiring of 100K edges to increase edge participation
$200K$	Enhance complexity	Rewiring of 200K edges to increase edge participation
$300K$	Enhance complexity	Rewiring of 300K edges to increase edge participation
$400K$	Enhance complexity	Rewiring of 400K edges to increase edge participation
$500K$	Enhance complexity	Rewiring of 500K edges to increase edge participation
$670K$	Enhance complexity	Rewiring of 670K edges to increase edge participation

In case of manipulations that *reduce complexity*, labeled *Order* M, we globally rewired the connectome across the complexity spectrum based on five simplified stochastic models of connectivity[76,44]. The total numbers of connections in these simplified connectomes were closely matched to the number of connections in the baseline connectome (relative difference less than 0.1 %). Self-connections (autapses) were not allowed. We used the following simplified models:

Order 1 Uniform connection probability between all pairs of pre- and post-synaptic neurons.

Order 2 Distance-dependent connection probability; the probability depends on the Euclidean distance between pre- and post-synaptic neurons (conceptually equivalent to the *Distance model* of Table 1, but with the technical difference that it was modelled by the sum of a proximal and distal exponential function[44]).

Order 3 Bipolar distance-dependent connection probability; same as *Order* 2 but with a distinction between pre-synaptic neurons above or below post-synaptic neurons (in axial direction perpendicular to the cortical layers).

Order 4 Offset-dependent connection probability; the probability depends on the axial and radial offsets between pre- and post-synaptic neurons (in a cylindrical coordinate system with axial direction perpendicular to the cortical layers).

Order 5 Position- and offset-dependent connection probability; same as *Order* 4 but with additional dependence on the absolute axial position (i.e., cortical depth) of the pre-synaptic neuron.

On the manipulations that *reduce reciprocity*, labeled $RC - M$, for selected reciprocal connections we removed one of the two directed edges at random. The reciprocal connections were selected from the spine[1] of maximal n-simplices as follows:

RC − 1 Reciprocal connections on the spine of maximal n-simplices for $n \geq 5$.

RC − 2 Reciprocal connections on the spine of maximal n-simplices for $n \geq 5$ and half (chosen at random) of the reciprocal connections that are on the spine of maximal n-simplices for $n = 4$ that are not in the spine of a higher dimensional simplex.

RC − 3 Reciprocal connections on the spine of maximal n-simplices for $n \geq 4$.

RC − 4 All reciprocal connections.

On the manipulations that *increase reciprocity*, labeled $RC + M$, we added reciprocal connections to existing edges such that the percentage of reciprocal connections is approximately multiplied by a chosen factor F_M across all maximal simplices. To do this, let G be the original connectome, for $1 \leq k$ let G_k be the subgraph on edges that are on simplices of dimension k (see Subsection 4.5) and let nrc_k be the number of reciprocal connections of G_k. We selected at random from the non-reciprocal edges of G_k, $F_M * nrc_k$ edges that we then turned into reciprocal connections. We did this for $F_1 = 2$ and $F_2 = 4$.

Finally, on the manipulations that *enhance complexity*, we use a custom algorithm presented by Reimann et al.[65] which exchanges edges in order to maximize the number of simplices an edge participates in. Briefly, for each edge in the network, we computed its *maximal dimension of edge participation*, which is the maximum dimension of simplices it participates in. For each node in the network and we computed its *n-node participation* for all dimensions n, which is the number of n-simplices the node participates in. Then we removed M edges at random with maximal

[1]Recall that the spine of an n-simplex is the set of edges $0 \to 1 \to 2 \to \ldots \to n$.

dimension edge participation smaller or equal to 3 (the middle dimension for this connectome) and added M new edges at random, where the probability of an edge $a \to b$ to be added was proportional to the participation of a and b in simplices above dimension 3. We performed these manipulations for $M = 100K, 200K, 300K, 400K, 500K,$ and $670K$ and labelled these manipulations as M.

4.9 Statistical metrics and tests

The one sample t-test was used to test if simplex counts and number of reciprocal connections of a given connectome are significantly larger/lower from the expected values on their corresponding random network controls. To compare the means of activity metrics (e.g., reliability, coupling coefficient, efficiency) across groups the Kruskal-Wallis-Test was used instead, since these distributions were not normal. In both cases, the sizes of the samples and p-values are reported in the figure captions.

To evaluate the non-symmetry of the uni-modal reliability distributions for the different manipulated connectomes, we computed their skewness using the Fisher-Pearson coefficient[77]. This parameter has a value of zero for normally distributed variables and a positive/negative value for unimodal distributions that are long tailed to the right/left.

4.10 Coupling coefficient

Following the ideas of Okun et al.[46], we calculated coupling coefficient of a neuron i as the Pearson correlation of its activity with the activity of the rest of the circuit and denoted it c_i. More precisely,

$$c_i = \frac{\sum_t (f_i(t) - \mu_i) \times (f_{\neq i}(t) - \mu_{\neq i})}{\sqrt{\sum_t (f_i(t) - \mu_i)^2 \times (f_{\neq i}(t) - \mu_{\neq i})^2}},$$

where f_i is either the neuron's deconvolved ΔF/F for MICrONS, or its binned (with 5 ms bin size) spike train for BBP, μ_i is f_i's mean, $f_{\neq i}(t)$ is the activity of the rest of the circuit i.e.,

$$f_{\neq i}(t) = \sum_{j \neq i} f_i(t),$$

and $\mu_{\neq i}$ is its mean. Note that when the number of neurons is large, $f_{\neq i}(t)$ is approximately the summed activity of all neurons. Thus, we used this value instead for the BBP data for an efficient implementation.

Furthermore, c_i values were normalized by z-scoring their differences from the average of 10 c_i values from surrogate datasets. Control datasets were created by randomly shuffling spike times, i.e., keeping single cell firing rates. We choose this *per-cell* normalization as opposed to *global* normalization (in which one would z-score c_i values by the mean and standard deviation of all neurons' surrogate c_is) as the global one resulted in a multimodal distribution for the BBP data (Fig. S9B1). The source of this artifact is that *in silico* we could record even the extremely low firing rate neurons, which lead to a low, secondary peak in the distribution (Fig. S9B2). On the other hand, for the MICrONS data set, both types of normalization returned virtually the same values (Fig. S9A).

4.11 Efficiency of a neighborhood

We define the *efficiency of a neighborhood* as the dimension of its activity normalized by the size of its active subneighborhood i.e., the number of nodes in it that fire[48–53] We determined

the dimensionality of the activity of the neighborhood using the correlation of the spike trains of its nodes. Thus, a neighborhood of maximum efficiency has maximally uncorrelated data and encodes more information, relative to its size.

Specifically, as for reliability, we first convolved the spike trains with a Gaussian kernel with $\sigma = 10\,\text{ms}$, to reduce the effect of noise and we call these the *spike signals*. A fraction of all the neurons (between $\sim 25\%$ and 30%) remain silent during a simulation. For each node v in G, let ℓ be the number of nodes in nodes in its neighborhood N_v that fire and let Corr_v denote the mean centered $\ell \times \ell$ matrix given by the pairwise Pearson correlation between the spike signals (see Fig 5B left for examples). We computed the singular values of Corr_v via singular value decomposition (SVD) and we denote by $\overrightarrow{ss_v} \in \mathbb{R}^\ell$ the cumulative vector of singular values normalized to lie between 0 and 1. The normalized dimension of the activity of N_v is the fraction of the ℓ nodes required to reach 0.9 of the cumulative spectrum i.e.,

$$\textit{Efficiency}(N_v) = \frac{n_{0.9}}{\ell},$$

where $n_{0.9}$ is the first position in $\overrightarrow{ss_v}$ at which 0.9 is reached (see Fig 5B right).

4.12 Classification

We used the neighborhood-based stimulus classification pipeline described by Conceição et al.[55] and Reimann et al.[54]. We briefly summarize their pipeline here, which we refer to as the *single selection procedure* and refer the reader to the original papers for details on their method.

First, we computed the neighborhood metrics listed in Table S1 for the neighborhoods of all excitatory neurons within the excitatory subcircuit and called these the *selection parameters*. Second, amongst all excitatory neurons, we picked 50 neighborhood centers either: at random or those that maximize or minimize a given *selection parameter*. Third, we featurized the activity of the selected neighborhoods as a response to each stimulus and reduced it to a set of time series that we call the *feature vectors*. The featurization procedure was done in two generic ways. On one hand, following Reimann et al.[54], in what we call the *PCA method*, we applied principal component analysis to the binned spike trains of the union of the selected neighborhoods with a bin size of 20 ms. On the other hand, following Conceição et al.[55], we considered the *active subnetworks* of each neighborhood, which are the subgraphs of the neighborhood on the nodes which are active in a given time window; for each stimulus presentation we considered two time windows of 25 ms each. This reduced the activity to a set of time series of graphs (one per stimulus), which we then transformed into feature vectors by calculating their value of one of the network metrics listed in Table S1. Finally, we classified these feature vectors using a Support Vector Machine (SVM) and computed the accuracy and cross-validation error.

Note that this pipeline provides for each (selection, featurization) parameter pair a classification accuracy obtained only from the activity of 50 neighborhoods rather than the full circuit. Moreover, the sub-selection of the activity data is done purely based on the structural properties of the circuit, in particular its neighborhood structure. However, in most cases the featurization procedure uses both the structural and activity data except for the PCA method which is based in the activity values only. In particular, the number or percentage of non-zero features within the feature vectors (which are ultimately the information that enters the classifier) depend not only on the size of the union of the neighborhoods but also on the (selection, feature) pairing and (see Fig. S11).

To test the effect of the neighborhood complexity on its efficiency as a classification readout we slightly modified the pipeline above and contrasted the results with the original. The modification was simple, and affected only the second step described above in which we replaced the single selection procedure by what we call the *double selection procedure*. Specifically, instead of

selecting the neighborhoods amongst all excitatory cells, we first selected those on the top 1% of reciprocal density (high complexity) or the bottom 1% (low complexity). We then selected neighborhoods from these either at random or according to a given selection parameter and followed the same steps described above.

4.13 Software and data availability

We used the following openly available software tools:

Connectome-Utilities[68] Utilities for running topological analyses on complex networks, available under https://github.com/BlueBrain/ConnectomeUtilities

Connectome-Analysis A library containing general functions to analyze connectomes, available under https://github.com/danielaegassan/connectome_analysis

TriDy/TopoSampling pipeline[55,54] Neighborhood-based stimulus classification pipelines, available under https://github.com/JasonPSmith/TriDy and https://github.com/BlueBrain/topological_sampling/tree/merge_samples, respectively

TriDy-tools Tools for creating new selection parameters and distributing featurization tasks for the TriDy pipeline, available under https://github.com/jlazovskis/TriDy-tools

Connectome-Manipulator[44] Framework for manipulations of connectomes in SONATA[45] format, will be made available upon journal submission.

Additionally, analysis code and processed data will be available on GitHub and Zenodo upon journal submission.

Acknowledgements

The authors would like to thank Guillaume Tauzin for his help with Poetry related questions in relation to `connalysis` and Elvis Boci for his help with visualizations.

Funding

This study was supported by funding to the Blue Brain Project, a research center of the École polytechnique fédérale de Lausanne (EPFL), from the Swiss government's ETH Board of the Swiss Federal Institutes of Technology.

Author contributions

- Conceptualization: D.E.S., C.P., J.L., J.P.S., M.W.R.
- Data curation: D.E.S., C.P., A.E., J.L., J.P.S., M.W.R.
- Formal analysis: D.E.S., C.P., A.E., J.L., J.P.S., M.W.R.
- Investigation: D.E.S., C.P., J.L., J.P.S.
- Methodology: D.E.S., C.P., A.E., J.L., M.S., J.P.S., K.H., R.L., M.W.R.
- Project administration: M.W.R.
- Software: D.E.S., C.P., A.E., J.L., M.S., J.P.S., M.W.R.

- Supervision: M.W.R.
- Validation: D.E.S., C.P., A.E., J.L., J.P.S., M.W.R.
- Visualization: D.E.S., C.P., A.E.
- Writing - original draft: D.E.S., C.P., A.E., M.W.R.
- Writing - review & editing: D.E.S., C.P., A.E., J.L., M.S., J.P.S., K.H., R.L., M.W.R.

Declaration of interests

The authors declare no competing interests.

References

[1] A Aldo Faisal, Luc PJ Selen, and Daniel M Wolpert. Noise in the nervous system. *Nature reviews neuroscience*, 9(4):292–303, 2008.

[2] Alfonso Renart and Christian K Machens. Variability in neural activity and behavior. *Current opinion in neurobiology*, 25:211–220, 2014.

[3] J Gerard G Borst. The low synaptic release probability in vivo. *Trends in neurosciences*, 33(6):259–266, 2010.

[4] Rodney J Douglas and Kevan AC Martin. Neuronal circuits of the neocortex. *Annu. Rev. Neurosci.*, 27:419–451, 2004.

[5] Kenneth D Miller. Canonical computations of cerebral cortex. *Current opinion in neurobiology*, 37:75–84, 2016.

[6] Max Nolte, Michael W Reimann, James G King, Henry Markram, and Eilif B Muller. Cortical reliability amid noise and chaos. *Nature communications*, 10(1):3792, 2019.

[7] Hsi-Ping Wang, Donald Spencer, Jean-Marc Fellous, and Terrence J Sejnowski. Synchrony of thalamocortical inputs maximizes cortical reliability. *Science*, 328(5974):106–109, 2010.

[8] Donald O. Hebb. *The Organization of Behavior; A Neuropsychological Theory*. John Wiley & Sons, Inc., New York, 1949. doi: 10.2307/1418888.

[9] Kenneth D Harris, Jozsef Csicsvari, Hajime Hirase, George Dragoi, and György Buzsáki. Organization of cell assemblies in the hippocampus. *Nature*, 424(6948):552–556, 2003.

[10] Vítor Lopes-dos Santos, Sidarta Ribeiro, and Adriano BL Tort. Detecting cell assemblies in large neuronal populations. *Journal of neuroscience methods*, 220(2):149–166, 2013.

[11] Juan A Gallego, Matthew G Perich, Lee E Miller, and Sara A Solla. Neural manifolds for the control of movement. *Neuron*, 94(5):978–984, 2017.

[12] Ashok Litwin-Kumar and Brent Doiron. Slow dynamics and high variability in balanced cortical networks with clustered connections. *Nature Neuroscience*, 15(11):1498–1505, November 2012. ISSN 1097-6256, 1546-1726. doi: 10.1038/nn.3220. URL https://www.nature.com/articles/nn.3220.

[13] Gustavo Deco and Etienne Hugues. Neural Network Mechanisms Underlying Stimulus Driven Variability Reduction. *PLoS Computational Biology*, 8(3):e1002395, March 2012. ISSN 1553-7358. doi: 10.1371/journal.pcbi.1002395. URL https://dx.plos.org/10.1371/journal.pcbi.1002395.

[14] Fereshteh Lagzi and Stefan Rotter. Dynamics of Competition between Subnetworks of Spiking Neuronal Networks in the Balanced State. *PLOS ONE*, 10(9):e0138947, September 2015. ISSN 1932-6203. doi: 10.1371/journal.pone.0138947. URL https://dx.plos.org/10.1371/journal.pone.0138947.

[15] Friedemann Zenke, Everton J. Agnes, and Wulfram Gerstner. Diverse synaptic plasticity mechanisms orchestrated to form and retrieve memories in spiking neural networks. *Nature Communications*, 6(6922), 2015. doi: 10.1038/ncomms7922.

[16] 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.

[17] 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.

[18] Emma K Towlson, Petra E Vértes, Sebastian E Ahnert, William R Schafer, and Edward T Bullmore. The rich club of the c. elegans neuronal connectome. *Journal of Neuroscience*, 33(15):6380–6387, 2013.

[19] Michael W Reimann, Max Nolte, Martina Scolamiero, Katharine Turner, Rodrigo Perin, Giuseppe Chindemi, Paweł Dłotko, Ran Levi, Kathryn Hess, and Henry Markram. Cliques of neurons bound into cavities provide a missing link between structure and function. *Frontiers in computational neuroscience*, 11:48, 2017.

[20] Michael W. Reimann, Sirio Bolaños-Puchet, Jean-Denis Courcol, Daniela Egas Santander, Alexis Arnaudon, Benoît Coste, Thomas Delemontex, Adrien Devresse, Hugo Dictus, Alexander Dietz, András Ecker, Cyrille Favreau, Gianluca Ficarelli, Mike Gevaert, Juan B. Hernando, Joni Herttuainen, James B. Isbister, Lida Kanari, Daniel Keller, James King, Pramod Kumbhar, Samuel Lapere, Jānis Lazovskis, Huanxiang Lu, Nicolas Ninin, Fernando Pereira, Judit Planas, Christoph Pokorny, Juan Luis Riquelme, Armando Romani, Ying Shi, Jason P. Smith, Vishal Sood, Mohit Srivastava, Werner Van Geit, Liesbeth Vanherpe, Matthias Wolf, Ran Levi, Kathryn Hess, Felix Schürmann, Eilif B. Muller, Srikanth Ramaswamy, and Henry Markram. Modeling and simulation of rat non-barrel somatosensory cortex. part i: Modeling anatomy. *bioRxiv*, 2022. doi: 10.1101/2022.08.11.503144. URL https://www.biorxiv.org/content/early/2022/08/15/2022.08.11.503144.

[21] György Buzsáki and Kenji Mizuseki. The log-dynamic brain: how skewed distributions affect network operations. *Nature Reviews Neuroscience*, 15(4):264–278, 2014.

[22] 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. *BioRxiv*, pages 2021–07, 2021.

[23] Henry Markram, Eilif Muller, Srikanth Ramaswamy, Michael W Reimann, et al. Reconstruction and simulation of neocortical microcircuitry. *Cell*, 163(2):456–492, 2015.

[24] Chi-Tin Shih, Yen-Jen Lin, Cheng-Te Wang, Ting-Yuan Wang, Chih-Chen Chen, Ta-Shun Su, Chung-Chuang Lo, and Ann-Shyn Chiang. Diverse community structures in the neuronal-level connectome of the drosophila brain. *Neuroinformatics*, 18:267–281, 2020.

[25] 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.

[26] Karan KH Manjunatha, Matteo Bruzzone, Giorgio Nicoletti, Samir Suweis, and Marco Dal Maschio. A network-based method for extracting the organization of brain-wide circuits from reconstructed connectome datasets. *bioRxiv*, pages 2023–05, 2023.

[27] András Ecker, Daniela Egas Santander, Sirio Bolaños-Puchet, James B. Isbister, and Michael W. Reimann. Cortical cell assemblies and their underlying connectivity: an in silico study. *PLoS Computational Biology*, 20(3):e1011891, 2024. doi: 10.1371/journal.pcbi.1011891.

[28] Ann E Sizemore, Chad Giusti, Ari Kahn, Jean M Vettel, Richard F Betzel, and Danielle S Bassett. Cliques and cavities in the human connectome. *Journal of computational neuroscience*, 44:115–145, 2018.

[29] 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.

[30] Miroslav Andjelković, Bosiljka Tadić, and Roderick Melnik. The topology of higher-order complexes associated with brain hubs in human connectomes. *Scientific reports*, 10(1):17320, 2020.

[31] Dinghua Shi, Zhifeng Chen, Xiang Sun, Qinghua Chen, Chuang Ma, Yang Lou, and Guanrong Chen. Computing cliques and cavities in networks. *Communications Physics*, 4(1):249, 2021.

[32] Jean-Gabriel Young, Giovanni Petri, Francesco Vaccarino, and Alice Patania. Construction of and efficient sampling from the simplicial configuration model. *Physical Review E*, 96(3): 032312, 2017.

[33] Florian Unger and Jonathan Krebs. Mcmc sampling of directed flag complexes with fixed undirected graphs. *arXiv preprint arXiv:2309.02938*, 2023.

[34] Daniel Witvliet, Ben Mulcahy, James K Mitchell, Yaron Meirovitch, Daniel R Berger, Yuelong Wu, Yufang Liu, Wan Xian Koh, Rajeev Parvathala, Douglas Holmyard, et al. Connectomes across development reveal principles of brain maturation. *Nature*, 596(7871):257–261, 2021.

[35] Michael Winding, Benjamin D Pedigo, Christopher L Barnes, Heather G Patsolic, Youngser Park, Tom Kazimiers, Akira Fushiki, Ingrid V Andrade, Avinash Khandelwal, Javier Valdes-Aleman, et al. The connectome of an insect brain. *Science*, 379(6636):eadd9330, 2023.

[36] Patric Hagmann, Maciej Kurant, Xavier Gigandet, Patrick Thiran, Van J Wedeen, Reto Meuli, and Jean-Philippe Thiran. Mapping human whole-brain structural networks with diffusion mri. *PloS one*, 2(7):e597, 2007.

[37] Travis A Jarrell, Yi Wang, Adam E Bloniarz, Christopher A Brittin, Meng Xu, J Nichol Thomson, Donna G Albertson, David H Hall, and Scott W Emmons. The connectome of a decision-making neural network. *science*, 337(6093):437–444, 2012.

[38] Lav R Varshney, Beth L Chen, Eric Paniagua, David H Hall, and Dmitri B Chklovskii. Structural properties of the caenorhabditis elegans neuronal network. *PLoS computational biology*, 7(2):e1001066, 2011.

[39] Matthew Kahle. Topology of random clique complexes. *Discrete mathematics*, 309(6):1658–1671, 2009.

[40] Carina Curto, Chad Giusti, Keler Marku, Eva Pastalkova, and Vladimir Itskov. Pairwise correlation graphs from hippocampal population activity have highly non-random, low-dimensional clique topology. *BMC neuroscience*, 14(1):1–2, 2013.

[41] 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.

[42] András Ecker, Daniela Egas Santander, Marwan Abdellah, Jorge Blanco Alonso, Sirio Bolaños-Puchet, Giuseppe Chindemi, James B Isbister, James Gonzalo King, Pramod Kumbhar, Ioannis Magkanaris, et al. Long-term plasticity induces sparse and specific synaptic changes in a biophysically detailed cortical model. *bioRxiv*, pages 2023–08, 2023.

[43] Eyal Gal et al. Rich cell-type-specific network topology in neocortical microcircuitry. *Nat. Neurosci.*, 20(7):1004–1013, 2017.

[44] Christoph Pokorny et al. A connectome manipulation framework for the systematic and reproducible study of structure–function relationships through simulations. *In preparation*, 2024.

[45] Kael Dai, Juan Hernando, Yazan N Billeh, Sergey L Gratiy, Judit Planas, Andrew P Davison, Salvador Dura-Bernal, Padraig Gleeson, Adrien Devresse, Benjamin K Dichter, et al. The sonata data format for efficient description of large-scale network models. *PLoS computational biology*, 16(2):e1007696, 2020.

[46] Michael Okun, Nicholas A Steinmetz, Lee Cossell, M Florencia Iacaruso, Ho Ko, Péter Barthó, Tirin Moore, Sonja B Hofer, Thomas D Mrsic-Flogel, Matteo Carandini, et al. Diverse coupling of neurons to populations in sensory cortex. *Nature*, 521(7553):511–515, 2015.

[47] Amiram Grinvald, Amos Arieli, Misha Tsodyks, and Tal Kenet. Neuronal assemblies: single cortical neurons are obedient members of a huge orchestra. *Biopolymers: Original Research on Biomolecules*, 68(3):422–436, 2003.

[48] John K Chapin and Miguel AL Nicolelis. Principal component analysis of neuronal ensemble activity reveals multidimensional somatosensory representations. *Journal of neuroscience methods*, 94(1):121–140, 1999.

[49] Eero P Simoncelli and Bruno A Olshausen. Natural image statistics and neural representation. *Annual review of neuroscience*, 24(1):1193–1216, 2001.

[50] Christian K Machens, Ranulfo Romo, and Carlos D Brody. Functional, but not anatomical, separation of "what" and "when" in prefrontal cortex. *Journal of Neuroscience*, 30(1):350–360, 2010.

[51] John P Cunningham and Byron M Yu. Dimensionality reduction for large-scale neural recordings. *Nature neuroscience*, 17(11):1500–1509, 2014.

[52] 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.

[53] Carsen Stringer, Marius Pachitariu, Nicholas Steinmetz, Matteo Carandini, and Kenneth D Harris. High-dimensional geometry of population responses in visual cortex. *Nature*, 571 (7765):361–365, 2019.

[54] Michael W Reimann, Henri Riihimäki, Jason P Smith, Jānis Lazovskis, Christoph Pokorny, and Ran Levi. Topology of synaptic connectivity constrains neuronal stimulus representation, predicting two complementary coding strategies. *PloS one*, 17(1):e0261702, 2022.

[55] Pedro Conceição, Dejan Govc, Jānis Lazovskis, Ran Levi, Henri Riihimäki, and Jason P Smith. An application of neighbourhoods in digraphs to the classification of binary dynamics. *Network Neuroscience*, 6(2):528–551, 2022.

[56] Andrew T Reid, Drew B Headley, Ravi D Mill, Ruben Sanchez-Romero, Lucina Q Uddin, Daniele Marinazzo, Daniel J Lurie, Pedro A Valdés-Sosa, Stephen José Hanson, Bharat B Biswal, et al. Advancing functional connectivity research from association to causation. *Nature neuroscience*, 22(11):1751–1760, 2019.

[57] Wolfgang Maass, Thomas Natschläger, and Henry Markram. Real-time computing without stable states: A new framework for neural computation based on perturbations. *Neural computation*, 14(11):2531–2560, 2002.

[58] Gergő Orbán, Pietro Berkes, József Fiser, and Máté Lengyel. Neural variability and sampling-based probabilistic representations in the visual cortex. *Neuron*, 92(2):530–543, 2016.

[59] Ruben S Van Bergen and Janneke FM Jehee. Probabilistic representation in human visual cortex reflects uncertainty in serial decisions. *Journal of Neuroscience*, 39(41):8164–8176, 2019.

[60] Katherine Morrison and Carina Curto. Predicting neural network dynamics via graphical analysis. In *Algebraic and Combinatorial Computational Biology*, pages 241–277. Elsevier, 2019.

[61] Caitlyn Parmelee, Samantha Moore, Katherine Morrison, and Carina Curto. Core motifs predict dynamic attractors in combinatorial threshold-linear networks. *PloS one*, 17(3): e0264456, 2022.

[62] Carina Curto, Christopher Langdon, and Katherine Morrison. Robust motifs of threshold-linear networks. *arXiv preprint arXiv:1902.10270*, 2019.

[63] Michael W Reimann, Anna-Lena Horlemann, Srikanth Ramaswamy, Eilif B Muller, and Henry Markram. Morphological diversity strongly constrains synaptic connectivity and plasticity. *Cerebral Cortex*, 27(9):4570–4585, 2017.

[64] Lida Kanari, Ying Shi, Alexis Arnaudon, Natalí Barros-Zulaica, Ruth Benavides-Piccione, Jay S Coggan, Javier DeFelipe, Kathryn Hess, Huib D Mansvelder, Eline J Mertens, et al. Of mice and men: topologically complex dendrites assemble uniquely human networks. *bioRxiv*, pages 2023–09, 2023.

[65] Michael W Reimann, Daniela Egas Santander, András Ecker, and Eilif B Muller. Specific inhibition and disinhibition in the higher-order structure of a cortical connectome. *bioRxiv*, pages 2023–12, 2023.

[66] Donald B Johnson. Finding all the elementary circuits of a directed graph. *SIAM Journal on Computing*, 4(1):77–84, 1975.

[67] Dmitri B Chklovskii, Thomas Schikorski, and Charles F Stevens. Wiring optimization in cortical circuits. *Neuron*, 34(3):341–347, 2002.

[68] Michael Reimann, András Ecker, A Sato, Kerem Kurban, and Olli Lupton. BlueBrain/ConnectomeUtilities: New features and examples: Reciprocal views. https://doi.org/10.5281/zenodo.10059227, October 2023.

[69] M. W. Reimann. Connectomics of (part of) the microns mm3 dataset. https://doi.org/10.5281/zenodo.8364070, September 2023.

[70] Michael Reimann, Daniela Egas Santander, and Christoph Pokorny. Connectivity matrix of the internal connectivity of an earlier version of an SSCX model, January 2024. URL https://doi.org/10.5281/zenodo.10599103.

[71] Pramod Kumbhar, Michael Hines, Jeremy Fouriaux, Aleksandr Ovcharenko, James King, Fabien Delalondre, and Felix Schürmann. CoreNEURON : An Optimized Compute Engine for the NEURON Simulator. *Frontiers in Neuroinformatics*, 13(63), 2019. doi: 10.3389/fninf.2019.00063.

[72] M. E. J. Newman. The Structure and Function of Complex Networks. *SIAM Review*, 45(2): 167–256, January 2003. ISSN 0036-1445, 1095-7200. doi: 10.1137/S003614450342480. URL http://epubs.siam.org/doi/10.1137/S003614450342480.

[73] Leonid V Kantorovich. Mathematical methods of organizing and planning production. *Management science*, 6(4):366–422, 1960.

[74] Susanne Schreiber, Jean-Marc Fellous, D Whitmer, Paul Tiesinga, and Terrence J Sejnowski. A new correlation-based measure of spike timing reliability. *Neurocomputing*, 52:925–931, 2003.

[75] Catherine S Cutts and Stephen J Eglen. Detecting pairwise correlations in spike trains: an objective comparison of methods and application to the study of retinal waves. *Journal of Neuroscience*, 34(43):14288–14303, 2014.

[76] Eyal Gal, Rodrigo Perin, Henry Markram, Michael London, and Idan Segev. Neuron geometry underlies universal network features in cortical microcircuits. *bioRxiv*, 2020.

[77] Daniel Zwillinger and Stephen Kokoska. *CRC standard probability and statistics tables and formulae*. Crc Press, 1999.

Supplementary material

Table S1: Feature and selection parameters. Each parameter is marked by S, F or S+F, denoting that it is used for selection only, featurization only, or selection and featurization, respectively. We have split the parameters into three categories: *Activity*, which considers only the activity of the neighborhoods and not their network structure, *Topology*, which consider a topological metric of the active subgraphs and *Spectral*, which consider a spectral property of the active subgraphs. Detailed definitions of all network the metrics can be found in Conceição et al.[55].

Abbreviation	Parameter name	Use	Type
pca	PCA of activity	F	Activity
ns	Neighborhood size	S+F	Activity
deg	Neuron degree	S+F	Topology
ideg	Neuron in-degree	S+F	Topology
odeg	Neuron out-degree	S+F	Topology
avdeg	Average degree	S	Topology
ec	Euler characteristic	S+F	Topology
nbc	Normalized Betti coefficient	S+F	Topology
fcc	Clustering coefficient	S+F	Topology
tcc	Transitive clustering coefficient	S+F	Topology
***i*simplex**	i-simplex count	S	Topology
***i*contain**	i-simplex containment	S	Topology
dc*i*	Simplex density	S+F	Topology
asg	Adjacency spectral gap	S+F	Spectral
asl	Adjacency spectral gap (low)	S+F	Spectral
asr	Adjacency spectral radius	S+F	Spectral
blsg	Bauer Laplacian spectral gap	S+F	Spectral
blsl	Bauer Laplacian spectral gap (low)	S+F	Spectral
blsr	Bauer Laplacian spectral radius	S+F	Spectral
blsRg	Reversed Bauer Laplacian spectral gaps	S+F	Spectral
blsRl	Reversed Bauer Laplacian spectral gap (low)	S+F	Spectral
blsRr	Reversed Bauer Laplacian spectral radius	S+F	Spectral
clsg	Chung Laplacian spectral gap	S+F	Spectral
clsh	Chung Laplacian spectral gap (high)	S+F	Spectral
clsr	Chung Laplacian spectral radius	S+F	Spectral
tpsg	Transition probability spectral gap	S+F	Spectral
tpsl	Transition probability spectral gap (low)	S+F	Spectral
tpsr	Transition probability spectral radius	S+F	Spectral
tpsRg	Reversed transition probability spectral gap	S+F	Spectral
tpsRl	Reversed transition probability spectral gap (low)	S+F	Spectral
tpsRr	Reversed transition probability spectral radius	S+F	Spectral

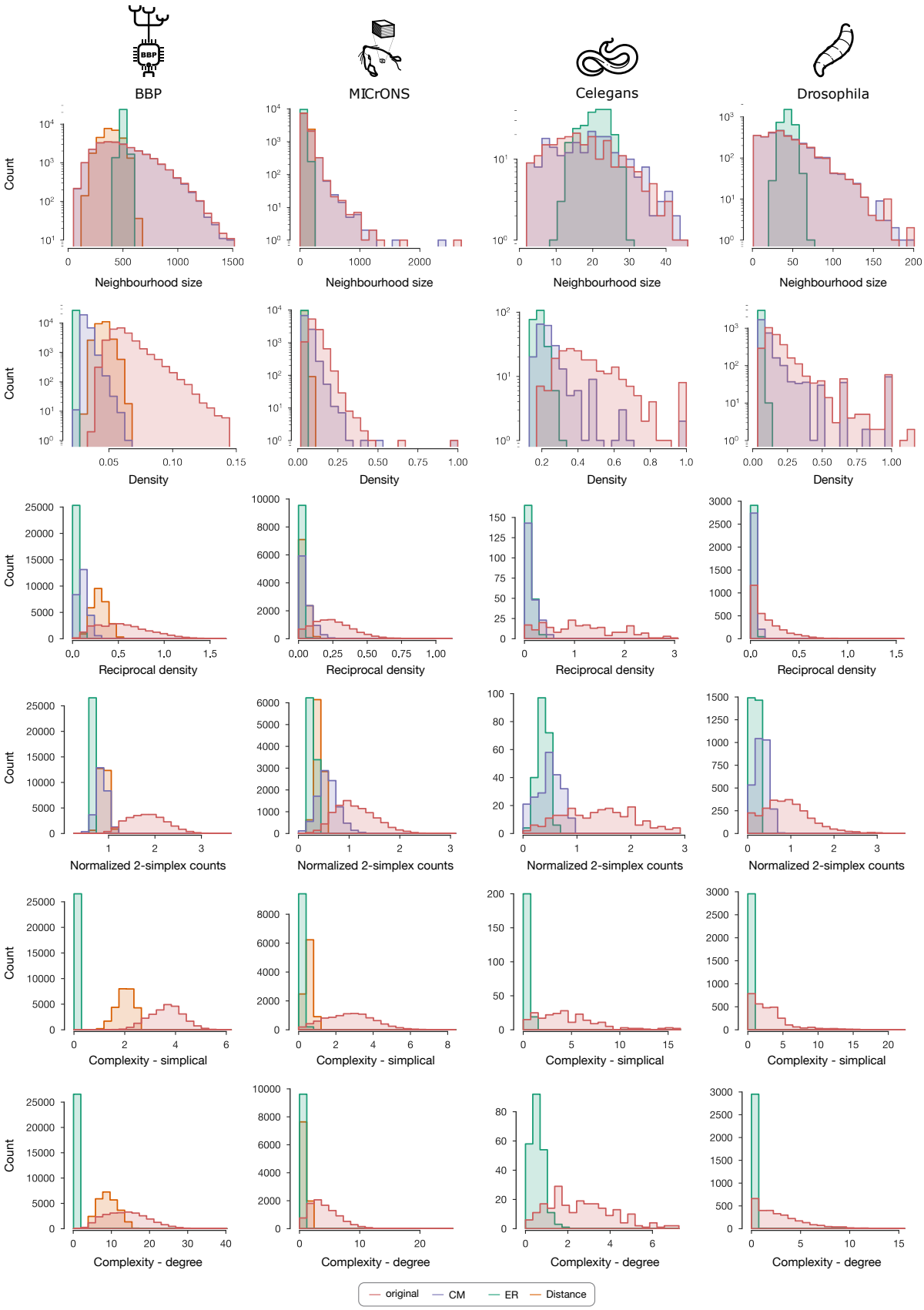

Figure S1: Network complexity promotes diversity of neighborhoods. Distribution of network metrics across neighborhoods for the original connectome and ER, CM and distance controls (when available). Row-by-row: Neighborhood size, density, reciprocal density, normalized 2-simplex counts, simplicial and degree complexity across connectomes. Normalization of 2-simplex counts was done by 1-simplex counts.

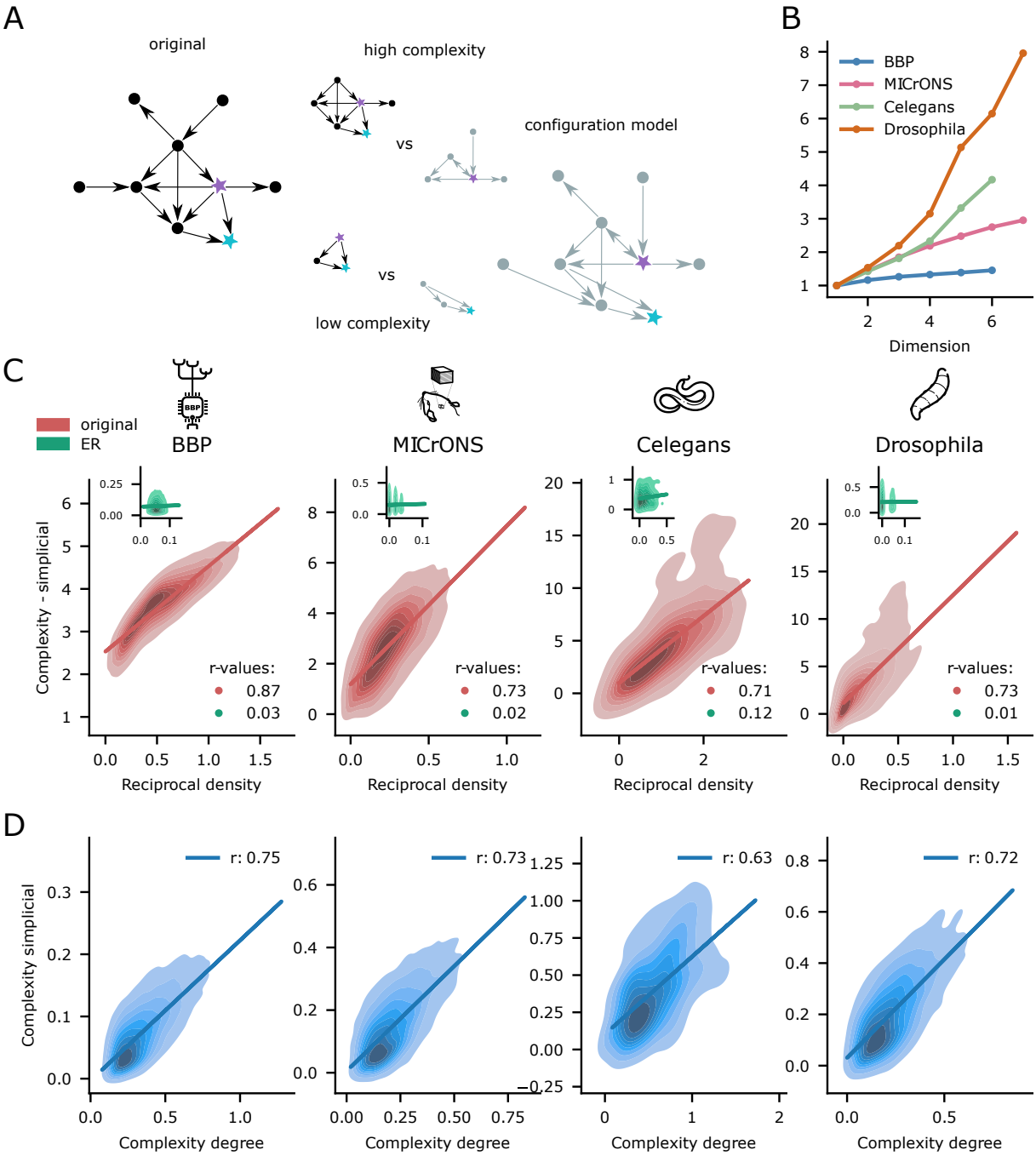

Figure S2: Neighborhood complexity (simplicial and degree). A: Toy example depicting the concept of neighborhood complexity. The neighborhoods of the purple/blue nodes have high/low complexity respectively. **B:** Average of neighborhood simplicial complexity on nodes that are on simplices of a given dimension(see Methods). **C:** In BNNs reciprocal density is strongly correlated with simplicial complexity. Inset, this is not the case for ER graphs. **D:** The metrics of degree complexity and simplicial complexity are strongly correlated.

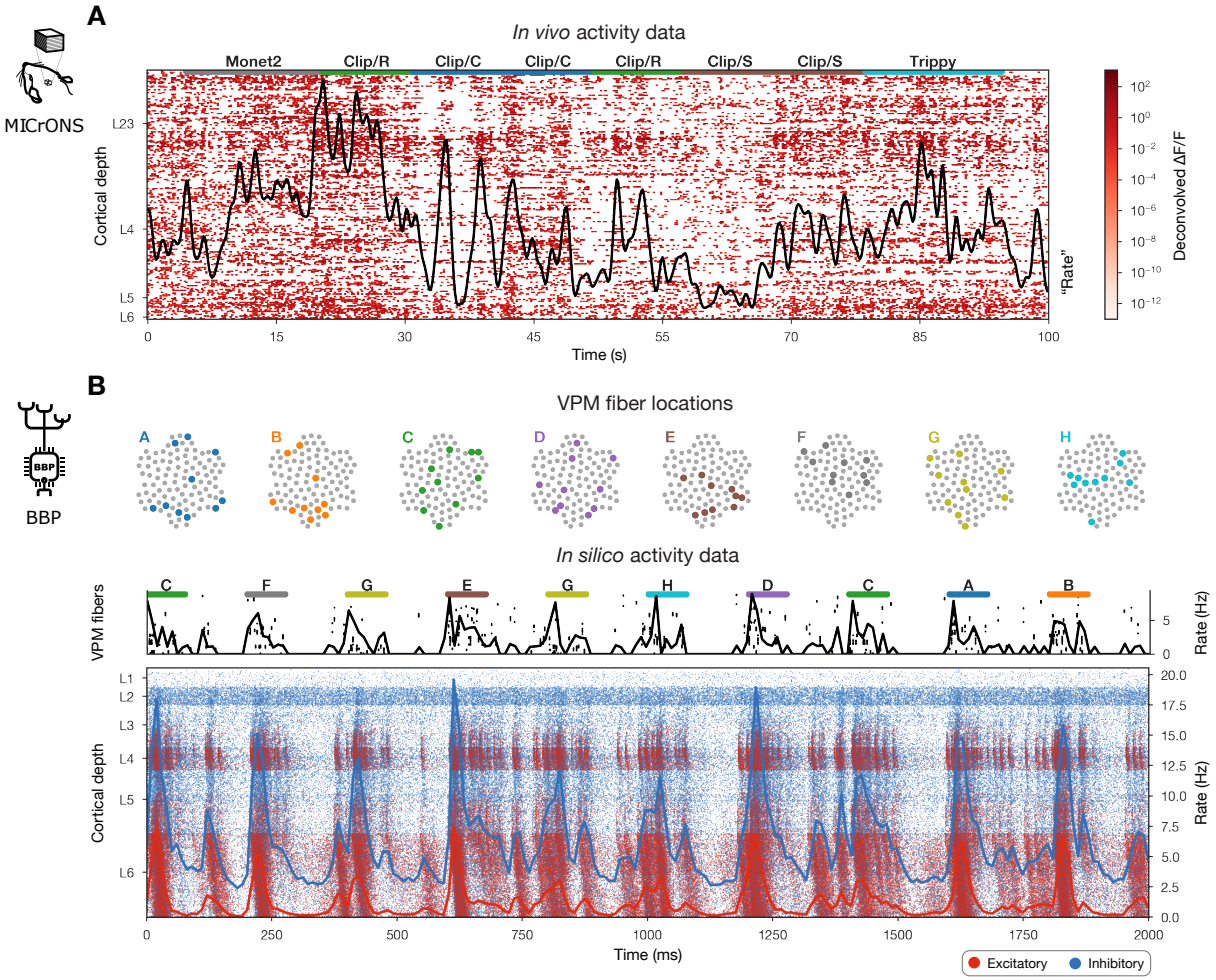

Figure S3: Activity datasets. Network activity and input stimuli on top for MICrONS (A) and BBP (B). **A:** MICrONS data shows deconvolved fluorescent traces, i.e., spike traces as extracted by the original authors[22]. It is 100 seconds from scan 7 from recording session 4. **B:** VPM input fiber bundles corresponding to the eight input patterns and input raster below. Raster plot of BBP network, excitatory cells in red, inhibitory ones (not analyzed) in blue.

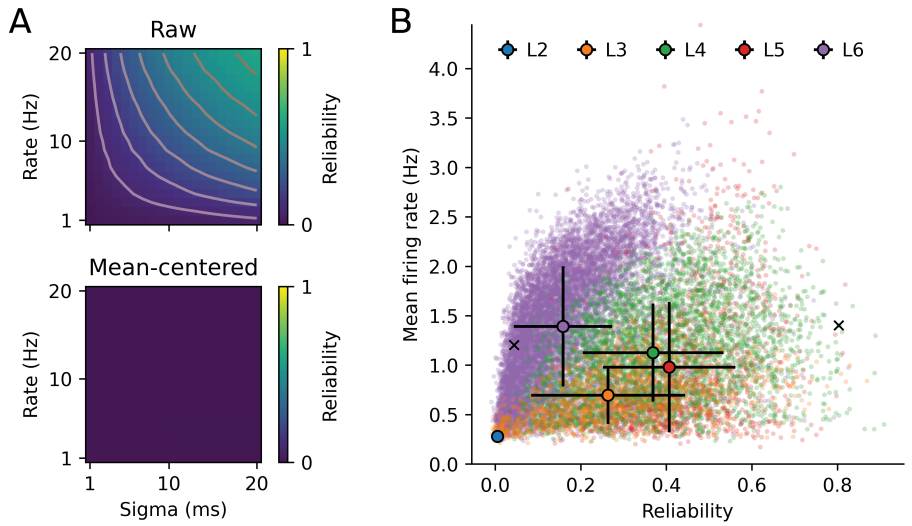

Figure S4: Reliability versus firing rates. A: Gaussian kernel reliability values of randomly generated Poisson spike trains with different firing rates and computed with different values of sigma. Without mean-centering of the data, there is a clear dependence of reliability on firing rates (upper panel) which disappears completely when mean-centering the data (lower panel). **B:** Mean-centered Gaussian kernel reliability values ($\sigma = 10\,\text{ms}$) versus mean firing rates of all excitatory neurons in BBP data, colored by layer. Error bars denote mean $\pm$ SD per layer. $\times$ markers indicate neurons with example spike trains shown in Fig. 3A.

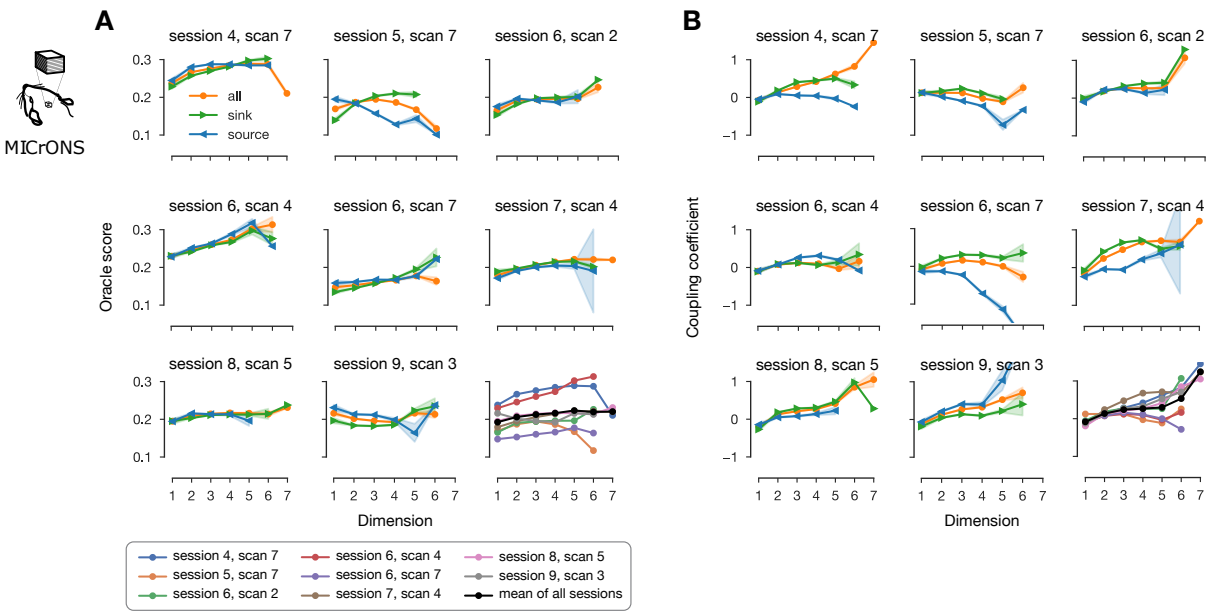

Figure S5: Per-scan activity metrics for MICrONS. Oracle score on panel **A** and coupling coefficient on panel **B** across simplices for the eight different recording sessions used (see Methods). The orange curves are the weighted averages of all neurons in that dimension, while the blue and green curves consider only the neurons in source or sink positions respectively. On the last subfigure, the curves for all nodes are pooled together across scans and the black line indicates the mean across sessions. (Mean across sessions is not the same as on the main figures, where the functional data was first z-scored and pooled together before running the analysis.) The same colors and legends as in A apply to B.

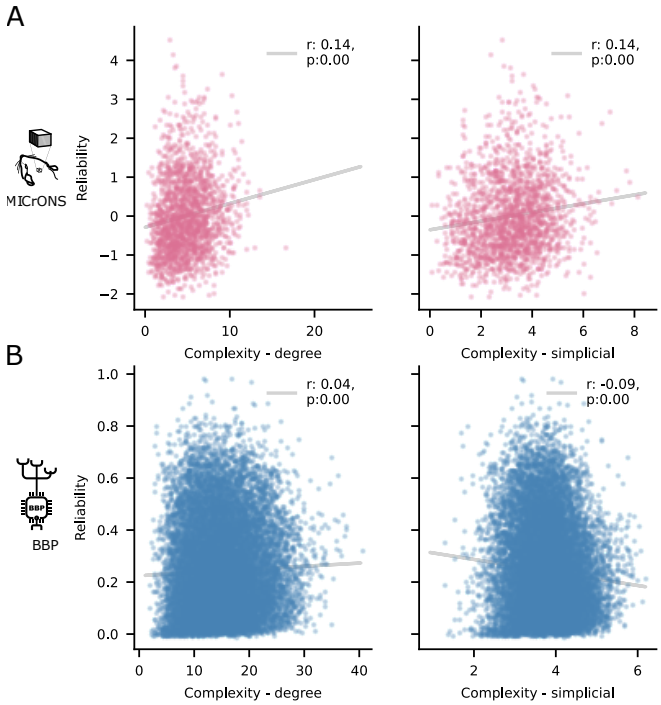

Figure S6: Reliability and neighborhood complexity. Negligible correlation is found between the neighborhood complexity and the reliability of its center, both for the degree and simplicial complexity metrics.

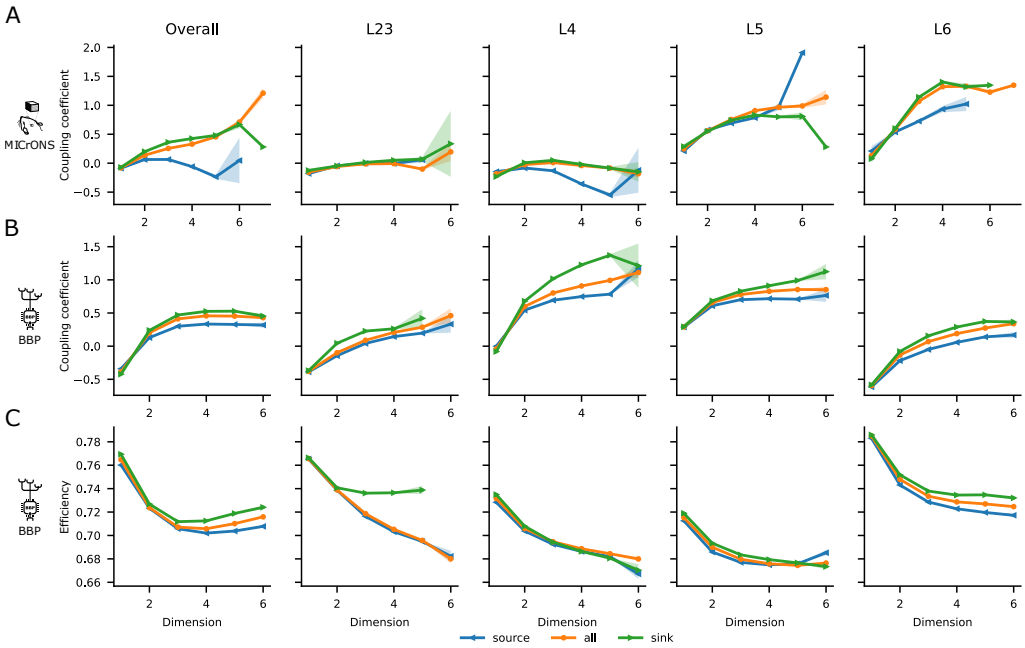

Figure S7: Coupling coefficient and efficiency across simplices split by layer. Coupling coefficient across simplices on panel **A** for MICrONS and **B** for BBP. On the left for all the neurons as in Figure 5A and on the right of it split across layers. **C:** As for **A** and **B** but for neighborhood efficiency (BBP data only). For all panels, the orange curves are the weighted averages of all neurons in that dimension, while the blue and green curves consider only the neurons in source or sink positions respectively.

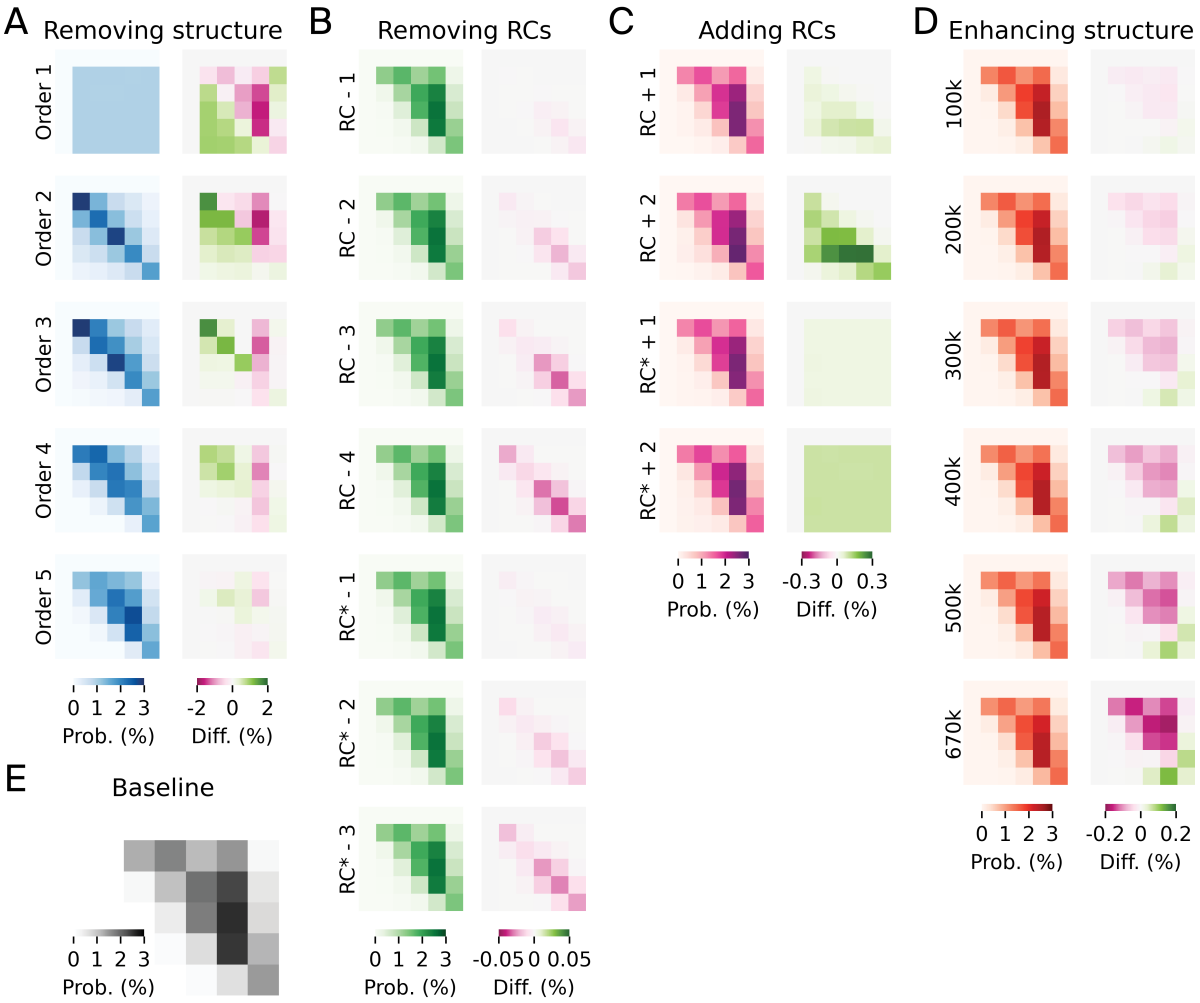

Figure S8: Structural validation of manipulated connectomes. For all manipulated connectomes shown in Fig. 4 and summarized in Table 2, the layer-wise connection probabilities when removing structure (**A**), removing reciprocal connections (**B**), adding reciprocal connections (**C**), and enhancing structure (**D**), and their differences to the baseline connectome (**E**) are illustrated. Asterisks indicate random controls. In the first column of each panel, the mean connection probabilities between excitatory neurons in each pre-synaptic (y axis) and post-synaptic (x axis) layer are shown. Note that layer 1 is empty. In the second column, the differences to the original connectome are highlighted. Note the different color scales at the bottom.

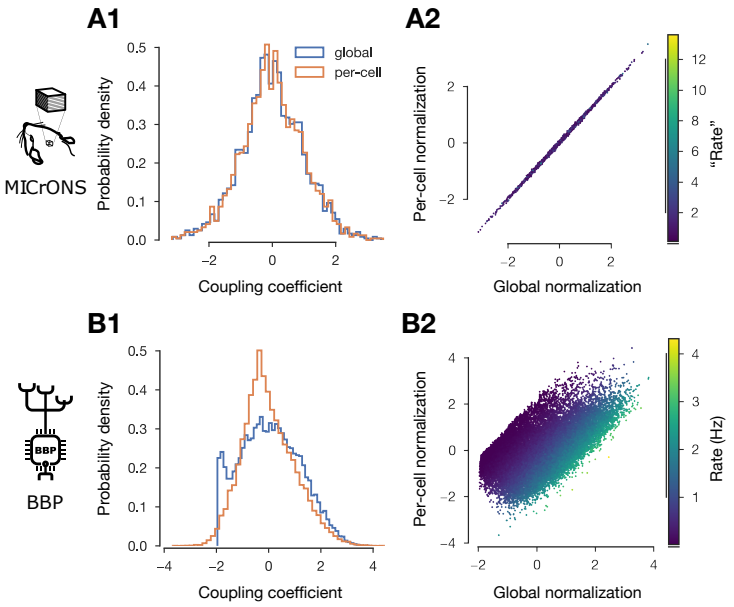

Figure S9: Normalization of raw coupling coefficients. A1 and **B1:** Global vs. stricter per-cell (keeping the firing rate, see Methods) normalization of coupling coefficients for MICrONS (A1) and BBP (B1). **A2** and **B2:** Correlation between global and per-cell normalized values and their link to firing rate (especially for BBP, where the two distributions differ.).

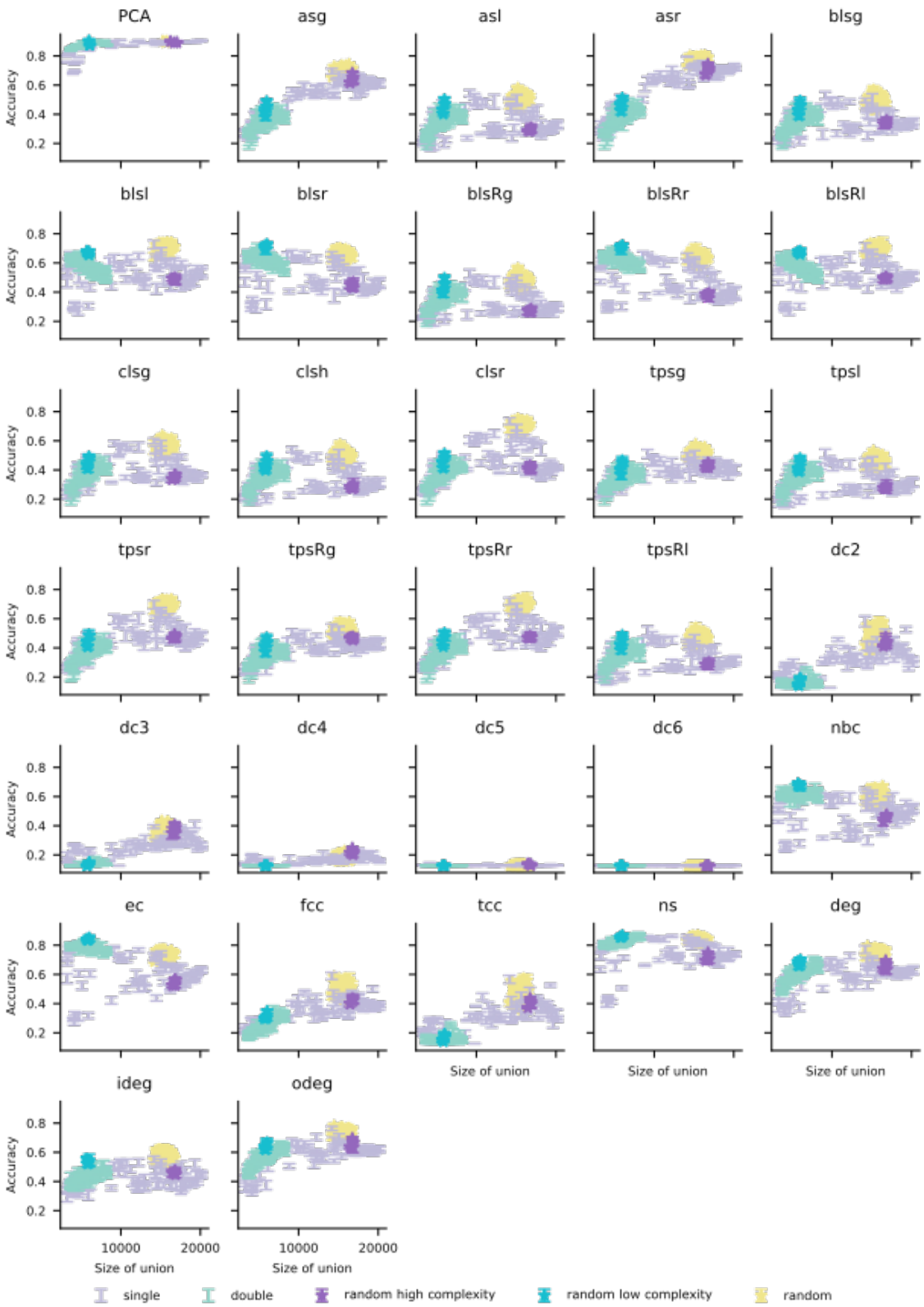

Figure S10: Classification accuracy for a given population size. Classification accuracy for all feature parameters. Each panel corresponds to a feature parameter, which is indicated in its title. Within each panel each bar corresponds to a selection of 50 neighborhoods selected either at random or maximizing/minimizing a given selection parameter. The x-axis of the bars corresponds to the size of the union of the 50 neighborhoods selected. Its center in the y-axis corresponds to the the accuracy for that (selection,feature) pair and the height of the bars is given by the cross validated error. Both selection and featurization parameters and their abbreviations are listed in Table S1.

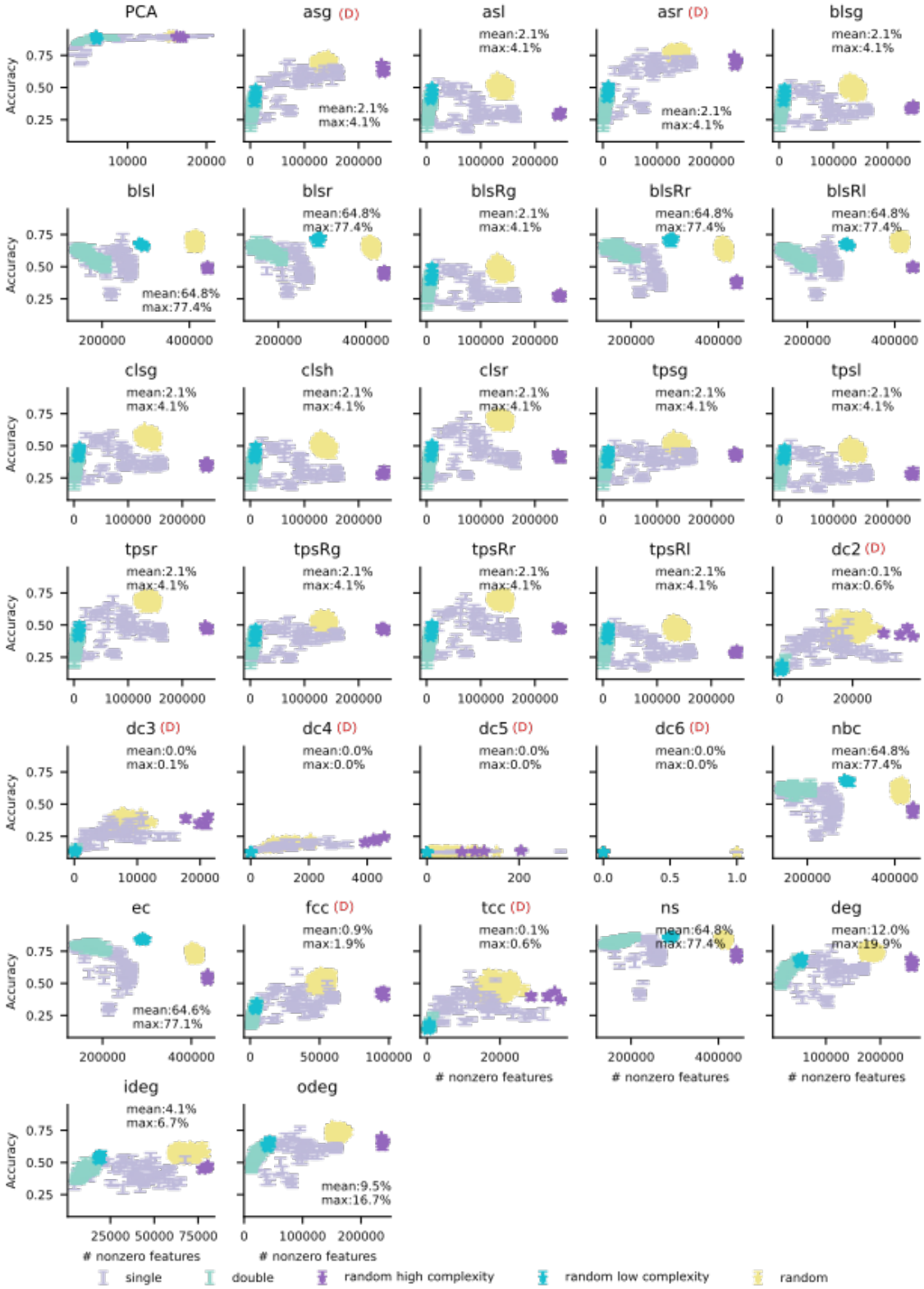

Figure S11: Classification accuracy versus number of nonzero features. Classification accuracy for all feature parameters as for Fig. S10 but where x-axis is given by the number of non-zero features in the featurization vector of that (selection,feature) pair. The legend indicates the mean and maximum percentage of nonzero entries in the feature vectors for the double classification procedure across all selection parameters. Degenerate featurization parameters (those for which most of the entries of their feature vectors for double selection are 0) have a (D) marking in their corresponding panels.