Geometry and dynamics of representations in a precisely balanced memory network related to olfactory cortex

Claire Meissner-Bernard[1], Friedemann Zenke[1,2], Rainer W. Friedrich[1,2*]

[1] Friedrich Miescher Institute for Biomedical Research, Maulbeerstrasse 66, 4058 Basel, Switzerland

[2] University of Basel, 4003 Basel, Switzerland

* Correspondence: rainer.friedrich@fmi.ch

Abstract

Biological memory networks are thought to store information in the synaptic connectivity between assemblies of neurons. Recent models suggest that these assemblies contain both excitatory and inhibitory neurons (E/I assemblies), resulting in co-tuning and precise balance of excitation and inhibition. To understand computational consequences of E/I assemblies under biologically realistic constraints we created a spiking network model based on experimental data from telencephalic area Dp of adult zebrafish, a precisely balanced recurrent network homologous to piriform cortex. We found that E/I assemblies stabilized firing rate distributions compared to networks with excitatory assemblies and global inhibition. Unlike classical memory models, networks with E/I assemblies did not show discrete attractor dynamics. Rather, responses to learned inputs were locally constrained onto manifolds that "focused" activity into neuronal subspaces. The covariance structure of these manifolds supported pattern classification when information was retrieved from selected neuronal subsets. Networks with E/I assemblies therefore transformed the geometry of neuronal coding space, resulting in continuous representations that reflected both relatedness of inputs and an individual's experience. Such continuous internal representations enable fast pattern classification, can support continual learning, and may provide a basis for higher-order learning and cognitive computations.

Introduction

Autoassociative memory establishes internal representations of specific inputs that may serve as a basis for higher brain functions including classification and prediction. Representational learning in autoassociative memory networks is thought to involve spike timing-dependent synaptic plasticity and potentially other mechanisms that enhance connectivity among assemblies of excitatory neurons (Hebb, 1949; Miehl et al., 2022; Ryan et al., 2015). Classical theories proposed that assemblies define discrete attractor states and map related inputs onto a common stable output pattern. Hence, neuronal assemblies are thought to establish internal representations, or memories, that classify inputs relative to previous experience via attractor dynamics (Hopfield, 1982; Kohonen, 1984). However, brain areas with memory functions such as the hippocampus or neocortex often exhibit dynamics that is atypical of attractor networks including irregular firing patterns, transient responses to inputs, and high trial-to-trial variability (Iurilli and Datta, 2017; Renart et al., 2010; Shadlen and Newsome, 1994).

Irregular, fluctuation-driven firing reminiscent of cortical activity emerges in recurrent networks when neurons receive strong excitatory (E) and inhibitory (I) synaptic input (Shadlen and Newsome, 1994; Van Vreeswijk and Sompolinsky, 1996). In such "balanced state" networks, enhanced connectivity among assemblies of E neurons is prone to generate runaway activity unless matched I connectivity establishes co-tuning of E and I inputs in individual neurons. The resulting state of "precise" synaptic balance stabilizes firing rates because inhomogeneities or fluctuations in excitation are tracked by correlated inhibition (Hennequin et al., 2017). E/I co-tuning has been observed experimentally in cortical brain areas (Bhatia et al., 2019; Froemke et al., 2007; Okun and Lampl, 2008; Rupprecht and Friedrich, 2018; Wehr and Zador, 2003) and emerged in simulations that included spike-timing dependent plasticity at I synapses (Lagzi and Fairhall, 2022; Litwin-Kumar and Doiron, 2014; Vogels et al., 2011; Zenke et al., 2015). In simulations, E/I co-tuning can be established by including I neurons in assemblies, resulting in "E/I assemblies" where I neurons track activity of E neurons (Barron et al., 2017; Lagzi and Fairhall, 2022; Mackwood et al., 2021). Exploring the structural basis of E/I co-tuning in biological networks is challenging because it requires the dense reconstruction of large neuronal circuits at synaptic resolution (Friedrich and Wanner, 2021).

Modeling studies started to investigate effects of E/I assemblies on network dynamics (Chenkov, 2017; Mackwood et al., 2021; Sadeh and Clopath, 2020a; Schulz et al., 2021) but the impact on neuronal computations in the brain remains unclear. Balanced state networks can exhibit a broad range of dynamical behaviors, including chaotic firing, transient responses and stable states (Festa et al., 2018; Hennequin et al., 2014; Litwin-Kumar and Doiron, 2012; Murphy and Miller, 2009; Rost et al., 2018; Roudi and Latham, 2007), implying that computational consequences of E/I assemblies depend on network parameters. We therefore examined effects of E/I assemblies on autoassociative memory in a spiking network model that was constrained by experimental data from telencephalic area Dp of adult zebrafish, which is homologous to piriform cortex (Mueller et al., 2011).

Dp and piriform cortex receive direct input from mitral cells in the olfactory bulb (OB) and have been proposed to function as autoassociative memory networks (Haberly, 2001; Wilson and Sullivan, 2011). Consistent with this hypothesis, manipulations of neuronal activity in piriform cortex affected olfactory memory (Meissner-Bernard et al., 2018; Sacco and Sacchetti, 2010). In both brain areas, odors evoke temporally patterned, distributed activity patterns (Blazing and Franks, 2020; Stettler and Axel, 2009; Yaksi et al., 2009) that are dominated by synaptic inputs from recurrent connections (Franks et al., 2011; Rupprecht and Friedrich, 2018) and modified by experience (Chapuis and Wilson, 2011; Frank et al., 2019; Jacobson et al., 2018; Pashkovski et al., 2020). Whole-cell voltage clamp recordings revealed that neurons in posterior Dp (pDp) received large E and I synaptic inputs during odor responses. These inputs were co-tuned in odor space and correlated on fast timescales, demonstrating that pDp enters a transient state of precise synaptic balance during odor stimulation (Rupprecht and Friedrich, 2018).

We found that network models of pDp without E/I co-tuning generated persistent attractor dynamics and exhibited a biologically unrealistic broadening of the firing rate distribution. Introducing E/I assemblies established E/I co-tuning, stabilized the firing rate distribution, and abolished persistent attractor states. In networks with E/I assemblies, population activity was locally constrained onto manifolds that represented learned and related inputs by "focusing" activity into neuronal subspaces. The covariance structure of manifolds supported pattern classification when information was retrieved from selected neuronal subsets. These results show that autoassociative memory networks constrained by biological data operate in a balanced regime where information is contained in the geometry of neuronal manifolds. Predictions derived from these analyses may be tested experimentally by measurements of neuronal population activity in zebrafish.

Results

A spiking network model based on pDp.

To analyze memory-related computational functions of E/I assemblies under biologically realistic constraints we created a spiking neural network model, pDp_{sim}, based on experimental data from pDp (Figure 1A) (Blumhagen et al., 2011; Rupprecht and Friedrich, 2018; Yaksi et al., 2009). pDp_{sim} comprised 4000 E neurons and 1000 I neurons, consistent with the estimated total number of 4000 – 10000 neurons in adult zebrafish pDp (unpublished observations). The network received afferent input from 1500 mitral cells in the OB with a mean spontaneous firing rate of 6 Hz (Friedrich and Laurent, 2004, 2001; Tabor and Friedrich, 2008). Odors were modeled by increasing the firing rates of 10% of mitral cells to a mean of 15 Hz and decreasing the firing rates of 5% of mitral cells to a mean of 2 Hz (Methods, Figure 1B). As a result, the mean activity increased only slightly while the variance of firing rates across the mitral cell population increased approximately 7-fold, consistent with experimental observations (Friedrich and Laurent, 2004; Wanner and Friedrich, 2020).

pDp_{sim} consisted of sparsely connected integrate-and-fire neurons with conductance-based synapses (connection probability ≤5%). Model parameters were taken from the literature when available and otherwise determined to reproduce experimentally accessible observables (Figure 1F-H; Methods). The mean firing rate was <0.1 Hz in the absence of stimulation and increased to ~1 Hz during odor presentation (Figure 1C, F) (Blumhagen et al., 2011; Rupprecht et al., 2021; Rupprecht and Friedrich, 2018). Population activity was odor-specific and activity patterns evoked by uncorrelated OB inputs remained uncorrelated in Dp (Figure 1H) (Yaksi et al., 2009). The synaptic conductance during odor presentation substantially exceeded the resting conductance and inputs from other E neurons contributed >80% of the excitatory synaptic conductance (Figure 1G). Hence, pDp_{sim} entered an inhibition-stabilized balanced state (Sadeh and Clopath, 2020b) during odor stimulation (Figure 1D, E) with recurrent input dominating over afferent input, as observed in Dp (Rupprecht and Friedrich, 2018). Shuffling spike times of inhibitory neurons resulted in runaway activity with a probability of ~80%, demonstrating that activity was indeed inhibition-stabilized. These results were robust against parameter variations (Methods). pDp_{sim} therefore reproduced key features of pDp.

Co-tuning and stability of networks with E/I assemblies

To create networks with defined neuronal assemblies we re-wired a small subset of the connections in randomly connected (*rand*) networks. An assembly representing a "learned" odor was generated by identifying the 100 E neurons that received the largest number of connections from the activated mitral cells representing this odor and increasing the probability of E-E connectivity among these neurons by the factor α (Figure 2A). The number of incoming connections per neuron was maintained by randomly eliminating connections from neurons outside the assembly. In each network, we created 15 assemblies representing uncorrelated odors. As a consequence, ~30% of E neurons were part of an assembly with few neurons participating in multiple assemblies. Odor-evoked activity within assemblies was higher than the population mean and increased with α (Figure 2B). When α reached a critical value of ~6, networks became unstable and generated runaway activity (Figure 2B).

We first set $\alpha = 5$ and scaled I-to-E connection weights uniformly by a factor χ ("*Scaled I*" networks) until population firing rates in response to learned odors were similar to firing rates in *rand* networks (Figure 2A, C, D, F). Under these conditions, activity within assemblies was still amplified substantially in comparison to the corresponding neurons in *rand* networks ("pseudo-assembly") whereas activity outside assemblies was substantially reduced (Figure 2E, G). Hence, non-specific scaling of inhibition resulted in a divergence of firing rates that exhausted the dynamic range of individual neurons in the population, implying that homeostatic global inhibition is insufficient to maintain a stable firing rate distribution. We further observed that neurons within activated assemblies produced regular spike trains (Supplementary Figure 2IIA, B), indicating that the balanced regime was no longer maintained.

In *rand* networks, correlations between E and I synaptic conductances in individual neurons were slightly above zero (Figure 2H), presumably as a consequence of stochastic

inhomogeneities in the connectivity (Pehlevan and Sompolinsky, 2014). In *Scaled I* networks, correlations remained near zero, indicating that E assemblies by themselves did not enhance E/I co-tuning (Figure 2H). *Scaled I* networks with structured E but random I connectivity can therefore not account for the stability, synaptic balance, and E/I co-tuning observed experimentally (Rupprecht and Friedrich, 2018).

We next created structured networks with more precise E/I balance by including I neurons within assemblies. We first selected the 25 I neurons that received the largest number of connections from the 100 E neurons of an assembly. The connectivity between these two sets of neurons was then enhanced by two procedures: (1) in "*Tuned I*" networks, the probability of I-to-E connections was increased by a factor β while E-to-I connections remained unchanged. (2) In "*Tuned E+I*" networks, the probability of I-to-E connections was increased by β and the probability of E-to-I connections was increased by γ (Figure 2A, Supplementary Figure 2IA). As for "*Scaled I*" networks, β and γ were adjusted to obtain mean population firing rates of ~1 Hz in response to learned odors (Figure 2F). The other observables used to constrain the *rand* networks remained unaffected (Supplementary Figure 2I B-D).

In *Tuned* networks, correlations between E and I conductances in individual neurons were significantly higher than in *rand* or *Scaled I* networks (Figure 2H. To further analyze E/I co-tuning we projected synaptic conductances of each neuron onto a line representing the E/I ratio expected in a balanced network ("balanced axis") and onto an orthogonal line ("counter-balanced axis"; Figure 2I). The ratio between the standard deviations along these axes has been used previously to quantify E/I co-tuning in experimental studies (Rupprecht and Friedrich, 2018). This ratio was close to 1 in *rand* and *Scaled I* networks but significantly higher in *Tuned I* and *Tuned E+I* networks (Figure 2I). Hence, *Tuned* networks exhibited significant co-tuning along the balanced axis, as observed in pDp (Rupprecht and Friedrich, 2018).

In *Tuned* networks, activity within assemblies was higher than the mean activity but substantially lower and more irregular than in *Scaled I* networks (Figure 2D, G; Supplementary Figure 2IIA,B). Unlike in *Scaled I* networks, mean firing rates evoked by novel odors were indistinguishable from those evoked by learned odors and from mean firing rates in *rand* networks (Figure 2F). Hence, E/I co-tuning prevented excessive amplification of activity in assemblies without affecting global network activity.

Effects of E/I assemblies on attractor dynamics

We next explored effects of assemblies on network dynamics. In *rand* networks, firing rates increased after stimulus onset and rapidly returned to a low baseline after stimulus offset. Correlations between activity patterns evoked by the same odor at different time points and in different trials were positive but substantially lower than unity, indicating high variability. Hence, *rand* networks showed transient and variable responses to input patterns, consistent with the typical behavior of generic balanced state networks (Shadlen and Newsome, 1994; Van Vreeswijk and Sompolinsky, 1996). *Scaled* networks responded to learned odors with

persistent firing of assembly neurons and high pattern correlations across trials and time, implying attractor dynamics (Hopfield, 1982; Khona and Fiete, 2022), whereas *Tuned* networks exhibited transient responses and modest pattern correlations similar to *rand* networks. Hence, *Tuned* networks did not exhibit stable attractor states, presumably because precise synaptic balance prevented strong recurrent amplification within E/I assemblies.

In classical memory networks, attractor dynamics mediates autoassociative pattern classification because noisy or corrupted versions of learned inputs converge onto a consistent output. Hence, classical attractor memory networks perform pattern completion, which may be assessed by different procedures. Completion of partial input patterns can be examined by stimulating subsets of E neurons in an assembly during baseline activity and testing for the recruitment of the remaining assembly neurons (Sadeh and Clopath, 2021; Vogels et al., 2011). We found that assemblies were recruited by partial inputs in all structured pDp_{sim} networks (*Scaled* and *Tuned*) without a significant increase in the overall population activity (Supplementary Figure 3A).

To assess pattern completion under more biologically realistic conditions, we morphed a novel odor into a learned odor (Figure 3A), or a learned odor into another learned odor. In *rand* networks, correlations between activity patterns across E neurons (output correlations) increased approximately linearly as a morphed odor approached a learned odor and remained lower than the corresponding pattern correlations in the OB (input correlations, Figure 3B and Supplementary Figure 3B). This is consistent with the absence of pattern completion in generic random networks (Babadi and Sompolinsky, 2014; Marr, 1969; Schaffer et al., 2018; Wiechert et al., 2010). *Scaled I* networks in contrast, showed typical signatures of pattern completion: output correlations increased abruptly as the learned odor was approached and eventually exceeded the corresponding input correlations (Figure 3B and Supplementary Figure 3B). In *Tuned* networks, output correlations changed approximately linearly and never exceeded input correlations, similar to observations in *rand* networks (Figure 3B and Supplementary Figure 3B). Similarly, firing rates of assembly neurons increased abruptly as the learned odor was approached in *Scaled I* networks but not in *Tuned* or *rand* networks (Figure 3C). Hence, networks with E/I assemblies did not perform pattern completion in response to naturalistic stimuli, consistent with the absence of stable attractor dynamics.

Geometry of activity patterns in networks with E/I assemblies

We next examined how E/I assemblies transform the geometry of neuronal representations, i.e their organization in a state space where each axis represents the activity of one neuron or one pattern of neural covariance (Chung and Abbott, 2021; Gallego et al., 2017; Langdon et al., 2023). To address this general question, we created an odor subspace and examined its transformation by pDp_{sim}. The subspace consisted of a set of OB activity patterns representing four uncorrelated pure odors, which were assigned to the corners of a square. Mixtures between the pure odors were represented by pixels inside the square. OB activity patterns representing mixtures were generated by selecting active mitral cells from each of the pure odors' patterns

with probabilities depending on the relative distances from the corners (Methods). Correlations between OB activity patterns representing pure odors and mixtures decreased approximately linearly as a function of distance in the subspace (Figure 4B). The odor subspace therefore represented a hypothetical olfactory environment with four odor sources at the corners of a square arena. Locations in the odor subspace were visualized by the color code depicted in Figure 4A.

To examine how pDp_{sim} transforms this odor subspace we projected time-averaged activity patterns onto the first two principal components (PCs). As expected, the distribution of OB activity patterns in PC space closely reflected the geometry of the square (Figure 4C). This geometry was largely maintained in the output of *rand* networks, consistent with the notion that random networks tend to preserve similarity relationships between input patterns (Babadi and Sompolinsky, 2014; Marr, 1969; Schaffer et al., 2018; Wiechert et al., 2010). We next examined outputs of *Scaled* or *Tuned* networks containing 15 assemblies, two of which were aligned with pure odors. The four odors delineating the odor subspace therefore consisted of two learned and two novel odors. In *Scaled I* networks, odor inputs were discretely distributed between three locations in state space representing the two learned odors and the residual odors, consistent with the expected attractor states (Figure 4D, E). *Tuned* networks, in contrast, generated continuous representations of the odor subspace (Figure 4D). The geometry of these representations was distorted in the vicinity of learned odors, which were further removed from most of the mixtures than novel odors. These geometric transformations were less obvious when activity patterns of *Tuned* networks were projected onto the first two PCs extracted from *rand* networks (Supplementary Figure 4A). Hence, E/I assemblies introduced local curvature into the coding space that partially separated learned from novel odors without disrupting the continuity of the subspace representation.

The curvature of the representation manifold in *Tuned* networks suggests that E/I assemblies confine activity along specific dimensions of the state space, indicating that activity was locally constrained onto manifolds. To test this hypothesis, we first quantified the dimensionality of odor-evoked activity by the participation ratio, a measure that depends on the eigenvalue distribution of the pairwise neuronal covariance matrix (Altan et al., 2021) (Methods). As expected, dimensionality was highest in *rand* networks and very low in *Scaled I* networks, reflecting the discrete attractor states (Figure 4F). In *Tuned* networks, dimensionality was high compared to *Scaled I* networks but lower than in *rand* networks (Figure 4F). The same trend was observed when we sampled data from a limited number of neurons to mimic experimental conditions (Supplementary Figure 4D). Furthermore, when restraining the analysis to activity evoked by novel odors and related mixtures, dimensionality was similar between *rand* and *Tuned* networks (Figure 4F). These observations, together with additional analyses of dimensionality (Supplementary Figure 4B, C), support the hypothesis that E/I assemblies locally constrain neuronal dynamics onto manifolds without establishing discrete attractor states. Generally, these observations are consistent with recent findings showing effects of specific circuit motifs on the dimensionality of neural activity (Dahmen et al., 2023; Recanatesi et al., 2019).

We further tested this hypothesis by examining the local geometry of activity patterns around representations of learned and novel odors. If E/I assemblies locally confine activity onto manifolds, small changes of input patterns should modify output patterns along preferred dimensions near representations of learned but not novel odors. To test this prediction, we selected sets of input patterns including each pure odor and the seven most closely related mixtures. We then quantified the variance of the projections of their corresponding output patterns onto the first 40 PCs (Figure 4G). This variance decreased only slightly as a function of PC rank for activity patterns related to novel odors, indicating that patterns varied similarly in all directions. For patterns related to learned odors, in contrast, the variance was substantially higher in the direction of the first few PCs, implying variation along preferred dimensions. In addition, we measured the distribution of angles between edges connecting activity patterns representing pure odors and their corresponding related mixtures in high-dimensional PC space (Figure 4H, inset; Methods; Schoonover et al., 2021). Angles were narrowly distributed around 1 rad in *rand* networks but smaller in the vicinity of learned patterns in *Tuned* networks (Figure 4 H). These observations further support the conclusion that E/I assemblies locally constrain neuronal dynamics onto manifolds.

Activity may be constrained non-isotropically by amplification along a subset of dimensions, by inhibition along other dimensions, or both. E neurons participating in E/I assemblies had large loadings on the first two PCs (Supplementary Figure 4 E-F) and responded to learned odors with increased firing rates as compared to the mean rates in *Tuned E+I* and *rand* networks. Firing rates of the remaining neurons, in contrast, were slightly lower than the corresponding mean rates in *rand* networks (Figure 2G). Consistent with these observations, the variance of activity projected onto the first few PCs was higher in *Tuned E+I* than in *rand* networks (Figure 4G) while the variance along higher-order PCs was lower (Figure 4G, inset). These observations indicate that activity manifolds are delineated both by amplification of activity along preferred directions and by suppression of activity along other dimensions.

Pattern classification by networks with E/I assemblies

The lack of stable attractor states raises the question how transformations of activity patterns by *Tuned* networks affect pattern classification. To quantify the association between an activity pattern and a class of patterns representing a pure odor we computed the Mahalanobis distance (d_M). This measure quantifies the distance between the pattern and the class center, normalized by the intra-class variability along the relevant direction. Hence, d_M is a measure for the discriminability of a given pattern from a given class. In bidirectional comparisons between patterns from different classes, the mean d_M may be asymmetric when the distributions of patterns within the classes are different.

We first quantified d_M between representations of pure odors based on activity patterns across 80 E neurons drawn from the corresponding (pseudo-) assemblies. d_M was asymmetrically increased in *Tuned E+I* networks as compared to *rand* networks. Large increases were observed for distances between patterns related to learned odors and reference

classes representing novel odors (Figure 5A, B). In the other direction, increases in d_M were smaller. Moreover, distances between patterns related to novel odors were almost unchanged (Figure 5B). Further analyses showed that increases in d_M in *Tuned E+I* networks involved both increases in the Euclidean distance between class centers and non-isotropic scaling of intra-class variability (Supplementary Figure 5). The geometric transformation of odor representations by E/I assemblies therefore facilitated pattern classification and particularly enhanced the discriminability of patterns representing learned odors.

To further analyze pattern classification, we performed multi-class quadratic discriminant analysis (QDA), an extension of linear discriminant analysis for classes with unequal variance. Using QDA, we determined the probabilities that an activity pattern evoked by a mixture is classified as being a member of each of four classes representing the pure odors, thus mimicking a 4-way forced choice odor classification task in the olfactory environment. The four classes were defined by the distributions of activity patterns evoked by the pure and closely related odors. We then examined the classification probability of patterns evoked by mixtures with respect to a given class as a function of the similarity between the mixture and the corresponding pure odor ("target") in the OB. As expected, the classification probability increased with similarity. Furthermore, in *Tuned E+I* networks, the classification probability of mixtures similar to a pure odor was significantly higher when the pure odor was learned (Figure 5C). Hence, E/I assemblies enhanced the classification of inputs related to learned patterns.

When neuronal subsets were randomly drawn not from assemblies but from the entire population, d_M was generally lower (Figure 5B). These results indicate that assembly neurons convey higher-than-average information about learned odors. Together, these observations imply that pDp_{sim} did not function as a categorical classifier but nonetheless supported the classification of learned odors, particularly when the readout focused on assembly neurons. Conceivably, classification may be further enhanced by optimizing the readout strategy, for example, by a learning-based approach. However, modeling biologically realistic readout mechanisms requires further experimental insights into the underlying circuitry.

Stability of networks with E/I assemblies against addition of memories

When networks successively learn multiple inputs over time, the divergence of firing rates and the risk of network instability is expected to increase as assemblies are accumulated. We surmised that *Tuned* networks should be more resistant against these risks than *Scaled* networks because activity is controlled more precisely, particularly when assemblies overlap. To test this hypothesis, we examined responses of *Tuned E+I* and *Scaled I* networks to an additional odor subspace where four of the six pairwise correlations between the pure odors were clearly positive (range, 0.27 – 0.44; Figure 6A). We then compared networks with 15 randomly created assemblies to networks with two additional assemblies aligned to two of the correlated pure odors. In *Scaled I* networks, creating two additional memories resulted in a substantial increase in firing rates, particularly in response to the learned and related odors. In *Tuned E+I* networks, in contrast, firing rates remained almost unchanged despite the increased memory load, and

representations of learned odors were well separated in PC space, despite the overlap between assemblies (Figure 6B, C). These observations are consistent with the assumption that precise balance in E/I assemblies protects networks against instabilities during continual learning, even when memories overlap. Furthermore, in this regime of higher pattern similarity, d_M was again increased upon learning, particularly between learned odors and reference classes representing other odors (not shown). E/I assemblies therefore consistently increased d_M in a directional manner under different conditions.

Discussion

A precisely balanced memory network constrained by pDp

Autoassociative memory networks map inputs onto output patterns representing learned information. Classical models proposed this mapping to be accomplished by discrete attractor states that are defined by assemblies of E neurons and stabilized by global homeostatic inhibition. However, as seen in *Scaled I* networks, global inhibition is insufficient to maintain a stable, biologically plausible firing rate distribution. This problem can be overcome by including I neurons in assemblies, which leads to precise synaptic balance. To explore network behavior in this regime under biologically relevant conditions we created a spiking network model constrained by experimental data from pDp. The resulting *Tuned* networks reproduced additional experimental observations that were not used as constraints including irregular firing patterns, lower output than input correlations, and the absence of persistent activity. Hence, pDp_{sim} recapitulated characteristic properties of a biological memory network with precise synaptic balance.

Neuronal dynamics and representations in precisely balanced memory networks

Simulated networks with global inhibition showed attractor dynamics and pattern completion, consistent with classical attractor memory. However, the distribution of firing rates broadened as connection density within assemblies increased, resulting in unrealistically high (low) rates inside (outside) assemblies and, consequently, in a loss of synaptic balance. Hence, global inhibition was insufficient to stabilize population activity against basic consequences of structured connectivity. In networks with E/I assemblies, in contrast, firing rates remained within a realistic range and the inhibition-stabilized balanced state was maintained. Such *Tuned* networks showed no discrete attractor states but transformed the geometry of the coding space by confining activity to continuous manifolds near representations of learned inputs.

Geometrical transformations in *Tuned* networks may be considered as intermediate between two extremes: (1) geometry-preserving transformations as, for example, performed by many random networks (Babadi and Sompolinsky, 2014; Marr, 1969; Schaffer et al., 2018; Wiechert et al., 2010), and (2) discrete maps as, for example, generated by discrete attractor networks (Freeman and Skarda, 1985; Hopfield, 1982; Khona and Fiete, 2022) (Figure 7). We found that

transformations became more discrete map-like when amplification within assemblies was increased and precision of synaptic balance was reduced. Likewise, decreasing amplification in assemblies of *Scaled* networks changed transformations towards the intermediate behavior, albeit with broader firing rate distributions than in *Tuned* networks (not shown). Hence, precise synaptic balance may be expected to generally favor intermediate over discrete transformations because this regime tends to linearize input-output functions (Baker et al., 2020; Denève and Machens, 2016).

E/I assemblies increased variability of activity patterns along preferred directions of state space and reduced their dimensionality in comparison to *rand* networks. Nonetheless, dimensionality remained high compared to *Scaled* networks with discrete attractor states. These observations indicate that geometric transformations in *Tuned* networks involved (1) a modest amplification of activity in one or a few directions aligned to the assembly, and (2) a modest reduction of activity in other directions. E/I assemblies therefore created a local curvature of coding space that "focused" activity in a subset of dimensions and, thus, stored information in the geometry of coding space.

As E/I assemblies were small relative to the total size of the E neuron population, stored information may be represented predominantly by small neuronal subsets. Consistent with this hypothesis, d_M was increased and the classification of learned inputs by QDA was enhanced when activity was read out from subsets of assembly neurons as compared to random neuronal subsets. Moreover, signatures of pattern completion were found in the activity of assemblies but not in global pattern correlations. The retrieval of information from networks with small E/I assemblies therefore depends on the selection of informative neurons for readout. Unlike in networks with global attractor states, signatures of memory storage may thus be difficult to detect experimentally without specific knowledge of assembly memberships.

Computational functions of networks with E/I assemblies

In theory, precisely balanced networks with E/I assemblies may support pattern classification despite high variability and the absence of discrete attractor states (Denève and Machens, 2016). Indeed, we found in *Tuned E+I networks* that input patterns were classified successfully by a generic classifier (QDA) based on selected neuronal subsets, particularly relative to learned inputs. Analyses based on the Mahalanobis distance d_M indicate that classification of learned inputs was enhanced by two effects: (1) local manifolds representing learned odors became more distant from representations of other odors due to a modest increase in firing rates within E/I assemblies, and (2) the concomitant increase in variability was not isotropic, remaining sufficiently low in directions that separated novel from learned patterns. Hence, information contained in the geometry of coding space can be retrieved by readout mechanisms aligned to activity manifolds. Efficient readout mechanisms may thus integrate activity primarily from assembly neurons, as mimicked in our QDA-based pattern classification. This notion is consistent with the finding that the integrated activity of E/I assemblies can be highly informative despite variable firing of individual neurons (Boerlin et

al., 2013; Denève et al., 2017; Denève and Machens, 2016). It will thus be interesting to explore how the readout of information from local manifolds could be further optimized.

Representations by local manifolds and discrete attractor states exhibit additional differences affecting neuronal computation: (1) *Tuned* networks do not mediate short-term memory functions based on persistent activity. Such networks may therefore support fast memoryless classification to interpret dynamical sensory inputs on a moment-to-moment basis. (2) The representation of learned inputs by small neuronal subsets, rather than global activity states, raises the possibility that multiple inputs can be classified simultaneously. (3) The stabilization of firing rate distributions by precise synaptic balance may prevent catastrophic network failures when memories are accumulated during continual learning. (4) The continuous nature of local manifolds indicates that information is not stored in the form of distinct items. Moreover, the coding space provides, in principle, a distance metric reflecting both relatedness in the feature space of sensory inputs and an individual's experience. Internal representations generated by precisely balanced memory networks may therefore provide a basis for higher-order learning and cognitive computations.

Balanced state networks with E/I assemblies as models for olfactory cortex

Piriform cortex and Dp have been proposed to function as attractor-based memory networks for odors. Consistent with this hypothesis, pattern completion and its modulation by learning has been observed in piriform cortex of rodents (Barnes et al., 2008; Chapuis and Wilson, 2011). However, odor-evoked firing patterns in piriform cortex and Dp are typically irregular, variable, transient and less reproducible than in the OB even after learning (Jacobson et al., 2018; Pashkovski et al., 2020; Schoonover et al., 2021; Yaksi et al., 2009), indicating that activity does not converge onto stable attractor states. Balanced networks with E/I assemblies, in contrast, are generally consistent with these experimental observations. Alternative models for pattern classification in the balanced state include networks endowed with adaptation, which respond to stimuli with an initial transient followed by a tonic non-balanced activity state (Wu and Zenke, 2021), or mechanisms related to "balanced amplification", which typically generate pronounced activity transients (Ahmadian and Miller, 2021; Murphy and Miller, 2009). However, it has not been explored whether these models can be adapted to reproduce characteristic features of Dp or piriform cortex.

Our results generate predictions to test the hypothesis that E/I assemblies establish local manifolds in Dp: (1) odor-evoked population activity should be constrained onto manifolds, particularly in response to learned odors. (2) Learning should increase the magnitude and asymmetry of d_M between odor representations. (3) Activity evoked by learned and related odors should exhibit lower dimensionality and more directional variability than activity evoked by novel odors. (4) Careful manipulations of inhibition may unmask assemblies by increasing amplification. These predictions may be addressed experimentally by large-scale measurements of odor-evoked activity after learning. The direct detection of E/I assemblies will ultimately require dense reconstructions of large neuronal networks at synaptic resolution. Given the small

size of Dp, this challenge may be addressed in zebrafish by connectomics approaches based on volume electron microscopy (Denk et al., 2012; Friedrich and Wanner, 2021; Kornfeld and Denk, 2018).

The hypothesis that memory networks contain E/I assemblies and operate in a state of precise synaptic balance can be derived from the basic assumptions that synaptic plasticity establishes assemblies and that firing rate distributions remain stable as network structure is modified by experience (Barron et al., 2017; Hennequin et al., 2017). Hence, *Tuned* networks based on Dp may also reproduce features of other recurrently connected brain areas such as hippocampus and neocortex, which also operate in a balanced state (Renart et al., 2010; Sadeh and Clopath, 2020b; Shadlen and Newsome, 1994) Future experiments may therefore explore representations of learned information by local manifolds also in cortical brain areas.

Methods

pDp spiking network model

pDp_{sim} consisted of 4000 excitatory (E) and 1000 inhibitory (I) neurons which were modeled as adaptive leaky integrate-and-fire units with conductance-based synapses of strength w.
A spike emitted by the presynaptic neuron y from population Y triggered an increase in the conductance g_{Yx} in the postsynaptic neuron x:

$$\tau_{syn,Y}\frac{dg_{Yx}}{dt} = -g_{Yx} + \sum_y w_{yx}\,\delta\left(t - t_{spike,y}\right) \text{ (equation 1)}$$

Neuron x received synaptic inputs from the olfactory bulb OB as well as from the different local neuronal populations P. Its membrane potential V_x evolved according to:

$$C_X\frac{dV_x}{dt} = g_{rest,X}\left(E_{rest,X} - V_x\right) + g_{OBx}(E_{exc} - V_x) + \sum_P g_{Px}(E_P - V_x) - z_x \text{ (equation 2)}$$

When the membrane potential reached a threshold V_{th}, the neuron emitted a spike and its membrane potential was reset to E_{rest} and clamped to this value during a refractory period τ_{ref}. Excitatory neurons were endowed with adaptation with the following dynamics (Brette and Gerstner, 2005):

$$\tau_a\frac{dz}{dt} = a(V_x - E) - z, \text{ with } z \to z + b \text{ after each spike (equation 3).}$$

In inhibitory neurons, z was set to 0 for simplicity.
The neuronal parameters of the model are summarized in Table 1. The values of the membrane time constant, resting conductance, inhibitory and excitatory reversal potential are in the range of experimentally measured values (Rupprecht and Friedrich, 2018 and Blumhagen et al., 2011). The remaining parameters were then fitted such as to fulfill two conditions (derived from unpublished experimental observations): (1) the neuron should not generate action potentials in response to a step current injection of duration 500 ms and amplitude 15 nA, and (2) the mean firing rate should be on the order of tens of Hz when the amplitude of the step current is 100 nA. Furthermore, the firing rates of inhibitory neurons should be higher than the firing rates of excitatory neurons, as observed experimentally (unpublished data).
The time constants of inhibitory and excitatory synapses ($\tau_{syn,I}$ and $\tau_{syn,E}$) were 10 ms and 30 ms, respectively. To verify that the behavior of pDp_{sim} was robust, we simulated 20 networks with different connection probabilities p_{YX} and synaptic strengths w_{YX} (Table 2). The connections between neurons were drawn from a Bernoulli distribution with a predefined $p_{YX} \leq 0.05$ (Zou, 2014). As a consequence, each neuron received the same number of input

connections. Care was also taken to ensure that the variation in the number of output connections was low across neurons.
The connection strengths w_{YX} were then fitted to reproduce experimental observations in pDp (five observables in total, see below and Figure 1). For this purpose, a lower and upper bound for w_{YX} were set such that the amplitude of single EPSPs and IPSPs was in the biologically plausible range of 0.2 to 2 mV. w_{OE} was then further constrained to maintain the odor-evoked, time-averaged g_{OE} in the range of experimental values (Rupprecht and Friedrich, 2018). Once w_{OE} was fixed, the lower bound of w_{EE} was increased to obtain a network rate >10 Hz in the absence of inhibition. A grid search was then used to refine the remaining w_{YX}.

Neuronal parameters	Symbol	Value	
		Excitatory neuron	Inhibitory neuron
Membrane time constant	τ_X	85 ms[1,2]	50 ms[2]
Resting conductance	g_{rest}	1.35 nS[1]	0.9 nS
Resting potential	E_{rest}	-60 mV[1,2]	-65 mV
Excitatory reversal potential	E_{exc}	0 mV[1]	0 mV
Inhibitory reversal potential	E_{inh}	-70 mV[1]	-70 mV
Spiking threshold	V_{th}	-38 mV[1]	-45 mV
Reset potential	E_{rest}	-60 mV	-65 mV
Refractory period	τ_{ref}	8 ms	8 ms
Adaptation time constant	τ_a	40 ms	/
Subthreshold adaptation	a	1 nS	/
Spike-triggered adaptation	b	10 pA	/

Table 1 Values of the neuronal parameters. The superscripts indicate the reference where the experimental measurements can be found. [1] Rupprecht and Friedrich, 2018 [2] Blumhagen et al., 2011

Network structure	**OB → E**	**OB → I**	**E → E**	**E → I**	**I → E**	**I → I**
A (n=5)	{0.02, 128}	{0.01, 68}	{0.05, 128}	{0.04, 68}	{0.05, 48}	{0.04, 25}
B (n=5)	{0.02, 128}	{0.01, 66}	{0.05, 128}	{0.04, 66}	{0.05, 45}	{0.04, 21}
C (n=5)	{0.02, 128}	{0.01, 68}	{0.05, 108}	{0.02, 80}	{0.05, 52}	{0.02, 31}
D (n=5)	{0.03, 95}	{0.03, 42}	{0.05, 128}	{0.04, 58}	{0.05, 59}	{0.04, 27}

Table 2 Values of the connectivity parameters of different networks {probability p_{YX} and synaptic strength w_{YX} in pS}

Olfactory bulb input

Each pDp neuron received external input from the OB, which consisted of 1500 mitral cells spontaneously active at 6 Hz. Odors were simulated by increasing the firing rate of 150 randomly selected mitral cells. Firing rates of these "activated" mitral cells were drawn from a discrete uniform distribution ranging from 8 to 32 Hz and their onset latencies were drawn from a discrete uniform distribution ranging from 0 to 200 ms. An additional 75 mitral cells were inhibited. Firing rates and latencies of these neurons were drawn from discrete uniform distributions ranging from 0 to 5 Hz and from 0 to 200 ms, respectively. After odor onset, firing rates decreased with a time constant of 1, 2 or 4 s (equally distributed). Spikes were generated from a Poisson distribution. Because all odors had almost identical firing patterns, the total OB input did not vary much across odors. In Figures 1 – 3, the odor set consisted of 10 novel and/or 10 learned odors, all of which were uncorrelated (pattern correlations near zero). Odors were presented for 2 seconds and separated by 1 second of baseline activity.
Olfactory subspaces comprised 121 OB activity patterns. Each pattern was represented by a pixel in a 11 x 11 square. The pixel at each vertex corresponded to one pure odor with 150 excited and 75 inhibited mitral cells as described above, and the remaining pixels corresponded to mixtures. The fraction of activated and inhibited mitral from a given pure odor decreased

with the distance from the corresponding vertex as shown in Table 3. The total number of activated and inhibited mitral cells at each location in the virtual square remained within the range of 150 ± 10% and 75 ± 10%, respectively. To generate activity patterns representing mixtures, activated mitral cells were sorted by onset latencies for each pure odor. At each location within the square and for each trial, mitral cells that remained activated in the mixture response were randomly selected from the pool of C mitral cells with the shortest latencies from each odor. C decreased with the distance from the vertices representing pure odors as shown in Table 4. The firing rate of each selected mitral cell varied ±1 Hz around its rate in response to the pure odor. The identity, but not the firing rate, of the activated mitral cells therefore changed gradually within the odor subspace. This procedure reflects the experimental observation that responses of zebrafish mitral cells to binary odor mixtures often resemble responses to one of the pure components (Tabor et al., 2004). We generated 8 different trajectories within the virtual square, each visiting all possible virtual odor locations for 1s. Each trajectory (trial) thus consisted of 121s of odor presentation, and trajectories were separated by 2 seconds of baseline activity. The dataset for analysis therefore comprised 968 activity patterns (8 trials x 121 odors).

100	90	80	70	60	50	40	30	20	10	0
90	90	80	70	60	50	40	30	15	5	0
80	80	70	60	50	40	35	25	15	5	0
70	70	60	50	50	35	30	20	15	5	0
60	60	50	50	40	30	25	20	15	5	0
50	50	40	35	30	25	25	15	10	5	0
40	40	35	30	25	25	15	10	10	5	0
30	30	25	20	20	15	10	10	5	5	0
20	15	15	15	15	10	10	5	5	5	0
10	5	5	5	5	5	5	5	5	0	0
0	0	0	0	0	0	0	0	0	0	0

Table 3 Percentage of cells selected from the 150 activated mitral cells defining one pure odor (vertex at top left) for uncorrelated pure odors.

100	100	100	90	80	70	60	50	40	30	0
100	100	100	90	80	70	60	50	35	25	0
100	100	90	60	70	60	55	45	35	25	0
90	90	80	70	70	55	50	40	35	25	0
80	80	70	70	60	50	45	40	35	25	0
70	70	60	55	50	45	45	35	30	25	0
60	60	55	50	45	45	35	30	30	25	0
50	50	45	40	40	35	30	30	25	25	0
40	35	35	35	35	30	30	25	25	25	0
30	25	25	25	25	25	25	25	25	0	0
0	0	0	0	0	0	0	0	0	0	0

Table 4 Percentage of activated cells available for selection for uncorrelated pure odors. C is obtained by multiplying the values by 1.5.

Assemblies

Unless noted otherwise, *Scaled* and *Tuned* networks contained 15 assembles ("memories"). An assembly representing a given odor contained the 100 E neurons that received the highest density of inputs from the corresponding active mitral cells. Hence, the size of assemblies was substantially smaller than the total population, consistent with the observation that only a minority of neurons in Dp or piriform cortex are activated during odor stimulation (Miura et al., 2012; Stettler and Axel, 2009; Yaksi et al., 2009) and upregulate *cfos* during olfactory learning (Meissner-Bernard et al., 2018). We then rewired assembly neurons: additional

connections were created between assembly neurons, and a matching number of existing connections between non-assembly and assembly neurons were eliminated. The number of input connections per neuron therefore remained unchanged. A new connection between two assembly neurons doubled the synaptic strength w_{EE} if it added to an existing connection. As a result of this rewiring, the connection probability within the assembly increased by a factor α relative to the baseline connection probability.
In *Scaled* networks, w_{IE} was increased globally by a constant factor χ. In *Tuned* networks, connections were modified between the 100 E neurons of an assembly and the 25 I neurons that were most densely connected to these E neurons, using the same procedure as for E-to-E connections. In *Tuned I* networks, only I-to-E connections were rewired, while in *Tuned E+I networks*, both I-to-E and E-to-I connections were rewired (Table 5). Whenever possible, networks with less than 15% change in population firing rates compared to the corresponding *rand* network were selected. In *Figure 6,* two additional assemblies were created in *Scaled* or *Tuned* networks without adjusting any parameters.

	Figure 2-6	**Supplementary Figure 2**
Scaled I	α=5, χ=1.4	α=3, χ=1.02; α=4, χ=1.1
Tuned I	α=5, β=18	α=3, β=11; α=4, β=12
Tuned E+I	α=5, β=4, γ=3	α=3, β=2.4, γ=1.4; α=4, β=3.2, γ=2

Table 5 Values of α, β, γ, χ used in the simulations. α: increase in E-E connection probability within assemblies. β: increase in I to E connection probability. γ: increase in E to I connection probability.

Observables
All variables were measured in the E population and time-averaged over the first 1.5 seconds of odor presentations, unless otherwise stated.
(1) The *firing rate* is the number of spikes in a time interval T divided by T.
(2) g_{OE} is the mean conductance change in E neurons triggered by spikes from the OB.
(3) g_{syn} is the total synaptic conductance change due to odor stimulation, calculated as the sum of g_{OE}, g_{EE} and g_{IE}. g_{EE} and g_{IE} are the conductance changes contributed by E synapses and I synapses, respectively.
(4) The *percentage of recurrent input* quantifies the average contribution of the recurrent excitatory input to the total excitatory input in E neurons. It was defined for each excitatory neuron as the ratio of the time-averaged g_{EE} to the time-averaged total excitatory conductance (g_{EE}+g_{OE}) multiplied by 100.
In (2,3,4), the time-averaged E and I synaptic conductances during the 500 ms before odor presentation were subtracted from the E and I conductances measured during odor presentation for each neuron.
(5) In addition, we required the Pearson correlation between activity patterns to be close to zero in response to uncorrelated inputs. The Pearson correlation between pairs of activity vectors composed of the firing rate of E neurons was averaged over all possible odor pairs.

Co-tuning
Co-tuning was quantified in 2 different ways: (1) For each neuron, we calculated the Pearson correlation between the time-averaged E and I conductances in response to 10 learned odors.
(2) As described in (Rupprecht and Friedrich, 2018), we projected observed pairs of E and I conductances onto a "balanced" and "counter-balanced" axis. The balanced axis was obtained by fitting a linear model without constant to the E and I conductances of 4000*10 neuron-learned odor pairs. The resulting model was a constant I/E ratio (~1.2) that defined a membrane potential close to spike threshold. The counter-balanced axis was orthogonal to the balanced

axis. For each neuron, synaptic conductances were projected onto these axes and their dispersions quantified by the standard deviations.

Characterization of population activity in state space

Principal Component Analysis (PCA) was applied to the OB activity patterns from the square subspace or to the corresponding activity patterns across E neurons in pDp_{sim} (8 x 121 = 968 patterns, each averaged over 1s).

The *participation ratio* PR provides an estimate of the maximal number of principal components (PC) required to recapitulate the observed neuronal activity. It is defined as $PR = \frac{(\sum_i \lambda_i)^2}{(\sum_i \lambda_i^2)}$, where λ_i are the eigenvalues obtained from PCA (variance along each PC).

For angular analyses we projected activity patterns onto the first 400 PCs, which was the number of PCs required to explain at least 75% of the variance in all networks. We measured the *angle* θ between the edges connecting the trial-averaged pattern $\mathbf{p}$ evoked by a pure odor to two patterns $\mathbf{s_y}$ and $\mathbf{s_z}$. $\mathbf{s_y}$ and $\mathbf{s_z}$ were trial-averaged patterns evoked by 2 out of the 7 odors that were most similar to the pure odor (21 angles in total). θ was defined as $\theta_{yz} = \cos^{-1}(\frac{(\mathbf{s_z}-\mathbf{p})\cdot(\mathbf{s_y}-\mathbf{p})}{\|\mathbf{s_z}-\mathbf{p}\|\|\mathbf{s_y}-\mathbf{p}\|})$. This metric is sensitive to non-uniform expansion and other non-linear transformations.

The *Mahalanobis distance* d_M is defined as $d_M(v, Q) = \sqrt{(v-\mu)^T S^{-1}(v-\mu)}$, where v is a vector representing an activity pattern, and Q a reference class consisting of a distribution of activity patterns with mean μ and covariance matrix S.

Classification

To assess the assignment of odor-evoked patterns to representations of pure odors in the odor subspace, we used *Quadratic Linear Discriminant* (QDA) analysis, a non-linear classifier that takes into account the separation and covariance patterns of different classes (Ghojogh and Crowley, 2019). The training set consisted of the population response to multiple trials of each of the 4 pure odors, averaged over the first and second half of the 1-s odor presentation. To increase the number of training data in each of the 4 odor classes, the training set also included population response to odors that were closely related to the pure odor (Pearson correlation between OB response patterns >0.6). Analyses were performed using subsets of 80 neurons (similar results were obtained using 50 – 100 neurons). These neurons were randomly selected (50 iterations) either from the (pseudo-) assemblies representing the pure odors (400 E neurons; pseudo-assemblies in *rand* networks and for novel odors) or from the entire population of E neurons. We verified that the data came from a Gaussian mixture model. The trained classifier was then applied to activity patterns evoked by the remaining odors of the subspace (correlation with pure odors < 0.6). Each pattern was then assigned to the class x that maximized the discriminant function

$\delta_k(x) = -\frac{1}{2}\log|S_k| - \frac{1}{2}(x-\mu_k)^T S_k^{-1}(x-\mu_k) + \log\pi_k$, where S_k is the covariance matrix of each odor class k (subsampling of neurons ensured invertibility) and π_k is the prior probability of class k. This discriminant function is closely related to d_M.

Simulations

Simulations were performed using Matlab and Python. Differential equations were solved using the forward Euler method and an integration time step of dt = 0.1 ms.

FIGURE LEGENDS

Figure 1. Spiking network model of pDp. A. Schematic of pDp_{sim}. OB, Olfactory bulb; E, excitatory; I, inhibitory neurons. **B**. Spike raster of a random subset of 50 mitral cells in the OB representing 2 odors. During odor stimulation, firing rates of 10% of mitral cells were increased and firing rates of 5% of mitral cells were decreased (baseline rate, 6 Hz). **C**. Spike raster of random subsets of 50 E and I neurons in response to 2 odors. **D**. Representative membrane potential trace (top) and excitatory (EPSC, black) and inhibitory (IPSC, red) currents (bottom) in one excitatory neuron in response to 2 odors. Purple trace shows net current (EPSC + IPSC). **E**. Odor-evoked inhibitory (red) and excitatory (black and blue) currents as measured in a hypothetical voltage clamp experiment (conductance multiplied by 70 mV, the absolute difference between holding potential and reversal potential; Rupprecht et al., 2018). Representative example of 1 network, averaged across neurons and odors. **F-H**. Measured values of the observables used to match pDp_{sim} to experimental data. Each dot represents one network (average over 10 odors); n = 20 networks. Pink shading shows the experimentally observed range of values. **F**. Baseline and odor-evoked population firing rate. **G.** Left: g_{OE} is the synaptic conductance in E neurons contributed by afferents from the OB during odor stimulation. Middle: g_{syn} is the total odor-evoked synaptic conductance. Right: % recurrent input quantifies the percentage of E input contributed by recurrent connections during odor stimulation. **H**. Correlation coefficient between odor-evoked activity patterns in Dp. The dotted line indicates the mean correlation between odor patterns in the OB.

Figure 2. **Neuronal assemblies (memories). A**. Schematic of assemblies. Each assembly contains the 100 E neurons that are most densely connected to the mitral cells activated by a given odor. Connection probability between these E neurons is increased by a factor α. In *Scaled I* networks, weights of all I-to-E synapses are increased by a factor χ. In *Tuned* networks, the 25 I neurons that are most densely connected to the 100 E neurons are included in each assembly. In *Tuned I* networks, the probability of I-to-E connections within the assembly is increased by a factor β. In *Tuned I+E* networks, probabilities of I-to-E and E-to-I connectivity within the assembly are increased by factors β and γ, respectively. n = 20 networks with 15 assemblies each were simulated for each group. **B.** Firing rates averaged over all E or I neurons (full lines) and over all assembly neurons (dashed lines) as a function of α (mean ± SD. across 20 networks). **C**. Mean E neurons firing rates of *Scaled* (left) and *Tuned* (right) networks in response to learned odors as a function of connection strength and probability, respectively. Squares depict parameters used in following figures unless stated otherwise. **D**. Spike raster plots showing responses of 50 E neurons to two odors in a *Scaled I* and the corresponding *Tuned E+I* network (same neurons and odors in the corresponding *rand* network are shown in Fig. 1C). **E**. Top: Mean firing rates in response to learned odors as a function of time, averaged across assembly or non-assembly neurons. Bottom: Correlation between activity patterns evoked by different trials of the same learned odor as a function of time. The pink bar indicates odor presentation. **F**. Mean firing rate in response to learned odors or novel odors. Each data point represents one network-odor pair (n=20 networks, 10 odors). **G**. Amplification within and outside (non-A.) assemblies, calculated as the ratio between mean firing rates in response to learned odors averaged across the same populations of neurons in a given structured network (*Scaled I*, *Tuned I*, or *Tuned E+I*) and the corresponding *rand* network. **H**. Quantification of co-tuning by the correlation between time-averaged E and I conductances in response to different odors, average across neurons (n = 20 networks). **I.** Quantification of co-tuning by the ratio of dispersion of joint conductances along balanced and counter-balanced axes (Methods). Each data point corresponds to one network (n = 20). Mean +/- SD.

Figure 3. Changes of output activity to gradual modifications of inputs. A. Morphing of a novel odor N into a learned odor L. Intermediate mixtures were generated by gradually decreasing/increasing the fractions of active mitral cells defining odors N/L. **B**. Pearson correlation between activity patterns evoked by the learned (full line) or novel (dotted line) odor and the intermediate odors in pDp as a function of the corresponding correlations in the OB. **C**. Firing rates in response to intermediate odors averaged across assembly neurons (learned odor, full line) or pseudo-assembly neurons (novel odor, dotted line) as a function of correlations between the OB activity patterns representing the intermediate odors and the OB activity pattern representing the learned odor. B, C: averages over 8 networks.

Figure 4. Geometry of odor representations in pDp_{sim}. A. Odor subspace delineated by 2 learned (L, M) and 2 novel (N, O) pure odors at the vertices of a square. Pixels within the square represent odor mixtures. **B.** Left: Pearson correlations between OB activity patterns defining pure odors. Right: Correlation between one pure odor (L; top left vertex) and all other odors. The odor from one vertex gradually morphs into the odor from another vertex. **C.** Projection of activity patterns in the OB onto the first 2 principal components (PCs). Colors represent patterns in the odor subspace shown in **A**. **D.** Projection of activity patterns in pDp in response to the odor subspace onto the first 2 PCs. Representative examples of different pDp networks. **E.** Density plot showing distribution of data points and demonstrating clustering at distinct locations in PC space for *Scaled I* networks. **F.** Quantification of dimensionality of neural activity by the participation ratio: activity evoked by novel odors and related mixtures (left), activity evoked by learned odors and related mixtures (center), and activity evoked by all stimuli (right). Each data point represents one network; dotted line represents the participation ratio of OB activity. **G.** Variance along the first 40 PCs extracted from activity patterns in *Rand* and *Tuned E+I* networks. Insets: variance along PCs 200 - 400. **H.** Angles between edges connecting a pure odor response and related versions thereof (see inset). The analysis was performed using the first 400 PCs, which explained >75% of the variance in all networks. n = 21 angles per pure odor in each of 8 networks (Methods). Similar results were obtained in the full-dimensional space.

Figure 5. Distance relationships and classification of odor representations. A. Activity patterns used as class distributions and vectors (B) or training and test sets (C). Same odor subspace as in Figure 4. **B**. Left: schematic illustration of Mahalanobis distance d_M. Right: d_M between one activity vector (v) and reference classes (Q) in *rand* and *Tuned E+I* networks. d_M was computed based on activity across subsets of 80 E neurons drawn from the four (pseudo-) assemblies with equal probability (top) or from the whole population (bottom). Note that d_M between patterns related to a learned odor and non-matching reference classes was higher in *Tuned E+I* networks, particularly when E neurons were drawn from assemblies. **C**. Pattern classification probability quantified by QDA. P_{Target} quantifies the probability that an activity pattern from the test set (odor mixtures, see **A**) is assigned to a target class from the training set (pure or closely related odor; see **A**). Left: classification probability as a function of the similarity (Pearson correlation) between the test and target odors in the OB (input patterns). Note enhanced classification probability for patterns evoked by odors similar to learned odors in *Tuned E+I* networks. Right: Classification probability for patterns similar to the training set (see **A**).

Figure 6. **Representation of correlated patterns and resilience against additional memories. A**. Subspace delineated by four positively correlated odors. Top: Correlations between pure odors. Bottom: Projection of OB activity patterns onto the first two PCs. **B**. Firing rates (top) and PC projection of output activity of a *Tuned E+I* network with 15 E/I assemblies that were not aligned to any of the four pure odors of the subspace. **C.** Firing rates (top) and PC

projection of output activity after creation of two additional assemblies representing two of the pure odors (Y and Z). Left: *Tuned E+I* network. Right: *Scaled I* network. Note that in the *Scaled I* network, but not in the *Tuned E+I* network, firing rates evoked by newly learned odors were increased and patterns evoked by these odors were not well separated in PC space.

Figure 7. Schematic of geometric transformations. A. Randomly connected networks tend to preserve the geometry of coding space. Such networks can support neuronal computations, e.g., by projecting activity patterns in a higher-dimensional coding space for pattern classification. **B**. We found that balanced networks with E/I assemblies transform the geometry of representations by locally restricting activity onto manifolds. These networks stored information about learned inputs while preserving continuity of the coding space. Such a geometry may support fast classification, continual learning and cognitive computations. Note that the true manifold geometry cannot be visualized appropriately in 2D because activity was "focused" in different subsets of dimensions at different locations of coding space. As a consequence, the dimensionality of activity remained substantial. **C**. Neuronal assemblies without precise balance established discrete attractor states, as observed in memory networks that store information as discrete items. Networks establishing locally defined activity manifolds (**B**) may thus be considered as intermediates between networks generating continuous representations without memories (**A**) and classical memory networks with discrete attractor dynamics (**C**).

Supplementary Figure Legends

Supplementary Figure 2I. Structured networks reproduce key features of Dp. A. Connection probability between classes of E and I neurons in a *rand*, a *Tuned I*, and a *Tuned E+I* network. **B.** EPSCs and IPSCs in a *Tuned I* network as observed in a hypothetical voltage clamp recording, averaged across neurons and odors. An equivalent plot for a *rand* network is shown in Figure 1E. **C**,**D**. Values of observables for *rand* networks (same as Figure 1G-H) and different structured networks (*Scaled I*, *Tuned I*, *Tuned E+I*). **E**. Network with increased connectivity between the E assembly neurons (α=5) and the 25 I neurons that are most densely connected to the mitral cells activated by a given odor. **F.** Mean firing rate of the network in **E** in response to learned odors as a function of connection probability. Selecting I neurons based on their afferent connectivity could not stabilize activity efficiently.

Supplementary Figure 2II. Structured networks: additional results. A. Raster plots showing responses of assembly (A) or non-assembly neurons to a learned odor. **B**. Coefficients of variation of the inter-spike interval (ISI) in assembly neurons. **C**. Correlation between activity patterns across E neurons evoked by the same novel odor in different trials as a function of time. Pink bar indicates odor presentation. Note that correlations in response to novel odors are similar across networks and different from responses to learned odors in *Scaled* and *Tuned* networks (Fig. 2E) **D**. Population firing rate in response to learned odors for networks with different probability of E-E connectivity within assemblies (α). **E**. Co-tuning (correlation between E and I currents in individual neurons) as a function of α.

Supplementary Figure 3. Pattern completion: additional results. A. Artificial reactivation of E assemblies. During 6 Hz baseline activity of the olfactory bulb, a subset of the assembly neurons was artificially reactivated by current injection (500 ms, 28 nA). Mean firing rates were quantified in the injected assembly neurons (i), in the remaining, non-injected assembly or pseudo-assembly neurons (ii), and in the non-assembly neurons (iii) as a function of time. The

orange bar indicates duration of current injection. **B**. Correlation between activity patterns across E neurons (output patterns) evoked by a series of input patterns representing a morph of one learned odor into another learned odor. Correlations between output patterns are plotted as a function of the correlation between the corresponding OB patterns.

Supplementary Figure 4. Transformations and dimensionality of activity patterns: additional results. A. Projection of activity patterns representing the odor subspace (Fig. 4A) onto the first 2 PCs of the corresponding *rand* networks (representative examples of one network each). **B**. Scree plot for the PCA results shown in Figure 4D. **C**. Error in the reconstruction of odor-evoked activity patterns as a function of the number of PCs. Euclidean distances between all pairs of activity patterns were calculated in the full-dimensional state space (Df) and in reduced-dimensional embedding spaces (Dl). The reconstruction error was defined as 1 – (correlation between Df and Dl). **D**. Participation ratio of the neural activity sampled from different numbers of neurons (50 iterations). **E**. Loadings of neurons on the first two PCs of a *rand* and a *Tuned E+I* network (Figure 4D). Each line represents one neuron. Neurons that are part of the assemblies representing the two learned odors are color-coded in magenta. **F**. For each network and PC, the 100 E neurons with the highest absolute loadings were selected and grouped into 3 categories: neurons part of the two assemblies representing the learned odors, neurons part of the two "pseudo-assemblies" representing the two novel odors, and the remaining neurons (non-assembly).

Supplementary Figure. 5. Further analyses of pattern distances. Differences in d_M between *rand* and *Tuned* networks may involve differences in the distance between class centers and/or differences in intra-class variability. To dissect the contributions of these effects we compared three distance measures:
(1) The Euclidean distance between class centers:

$$d_E = \sqrt{(\mu_V - \mu_Q)^T (\mu_V - \mu_Q)}$$

(2) The mean Euclidean distance between patterns of one class and the center of another class ($d_{\hat{E}}$):

$$d_{\hat{E}} = \frac{1}{n} \sum_{i=1..n} \sqrt{(v_i - \mu_Q)^T (v_i - \mu_Q)}$$

(3) The Mahalanobis distance d_M:

$$d_M = \frac{1}{n} \sum_{i=1..n} \sqrt{(v_i - \mu_Q)^T S^{-1} (v_i - \mu_Q)}$$

Here, μ_V and μ_Q are the centers (average patterns) of classes V and Q, respectively; v is a vector from V = $[v_1, v_2, \ldots, v_n]$, and S^{-1} is the inverse of the covariance matrix of the neuronal population within reference class Q. Note that $d_{\hat{E}}$ corresponds to d_M without normalization by variability (covariance).
A. Analysis of d_E. Left: d_E in *rand* and *Tuned E+I* networks based on 80 E neurons drawn from (pseudo-) assemblies. Right: same based on 80 E neurons drawn from the entire population. Note that d_E between learned and other odors was increased in *Tuned E+I* networks as compared to *rand* networks, particularly when neurons were drawn from assemblies. **B**. Equivalent plots for $d_{\hat{E}}$. Note that distances were increased nearly symmetrically, similar to d_E. **C**. Equivalent plots for d_M (same plots as in Fig. 5B). Note that d_M was increased asymmetrically. These observations show that the changes in d_M relative to *rand* networks involved an increase in the distance between class centers (d_E) and a non-isotropic change in intra-class variability (comparison between $d_{\hat{E}}$ and d_M). These effects were prominent when E neurons were drawn from assemblies. An important contribution to the increase in d_M in the

direction from learned odors to reference classes representing novel odors was made by the increased distance between class centers. In the other direction, d_M was smaller, implying that variability in the reference class was higher. Nonetheless, variability in the relevant direction did not fully counteract the increased distance between class centers in *Tuned E+I* networks. As a consequence, d_M was still increased slightly relative to the corresponding *rand* networks. Most of these effects were still observed, albeit weakly, when E neurons were drawn from the whole population.

Acknowledgements
We thank the Friedrich lab for insightful discussions. This work was supported by the Novartis Research Foundation, by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement no. 742576), and by the Swiss National Science Foundation (grants no. 31003A_172925/1, PCEFP3_202981).

References

Ahmadian Y, Miller KD. 2021. What is the dynamical regime of cerebral cortex? *Neuron* **109**:3373–3391. doi:10.1016/j.neuron.2021.07.031

Altan E, Solla SA, Miller LE, Perreault EJ. 2021. Estimating the dimensionality of the manifold underlying multi-electrode neural recordings. *Plos Comput Biol* **17**:e1008591. doi:10.1371/journal.pcbi.1008591

Babadi B, Sompolinsky H. 2014. Sparseness and Expansion in Sensory Representations. *Neuron* **83**:1213–1226. doi:10.1016/j.neuron.2014.07.035

Baker C, Zhu V, Rosenbaum R. 2020. Nonlinear stimulus representations in neural circuits with approximate excitatory-inhibitory balance. *Plos Comput Biol* **16**:e1008192. doi:10.1371/journal.pcbi.1008192

Barnes DC, Hofacer RD, Zaman AR, Rennaker RL, Wilson DA. 2008. Olfactory perceptual stability and discrimination. *Nat Neurosci* **11**:1378–1380. doi:10.1038/nn.2217

Barron HC, Vogels TP, Behrens TE, Ramaswami M. 2017. Inhibitory engrams in perception and memory. *Proc National Acad Sci* **114**:6666–6674. doi:10.1073/pnas.1701812114

Bhatia A, Moza S, Bhalla US. 2019. Precise excitation-inhibition balance controls gain and timing in the hippocampus. *eLife* 1–29. doi:10.7554/elife.43415.001

Blazing RM, Franks KM. 2020. ScienceDirect Odor coding in piriform cortex: mechanistic insights into distributed coding. *Elsevier Ltd* 1–7. doi:10.1016/j.conb.2020.03.001

Blumhagen F, Zhu P, Shum J, Schärer Y-PZ, Yaksi E, Deisseroth K, Friedrich RW. 2011. Neuronal filtering of multiplexed odour representations. *Nature* **479**:493–498. doi:10.1038/nature10633

Boerlin M, Machens CK, Denève S. 2013. Predictive Coding of Dynamical Variables in Balanced Spiking Networks. *PLoS Comput Biol* **9**:e1003258. doi:10.1371/journal.pcbi.1003258

Brette R, Gerstner W. 2005. Adaptive Exponential Integrate-and-Fire Model as an Effective Description of Neuronal Activity. *J Neurophysiol* 1–6. doi:10.1152/jn.00686.2005

Chapuis J, Wilson DA. 2011. Bidirectional plasticity of cortical pattern recognition and behavioral sensory acuity. *Nat Neurosci* **15**:155–161. doi:10.1038/nn.2966

Chenkov HSRKN. 2017. Memory replay in balanced recurrent networks 1–36. doi:10.1371/journal.pcbi.1005359

Chung S, Abbott LF. 2021. Neural population geometry: An approach for understanding biological and artificial neural networks. *Curr Opin Neurobiol* **70**:137–144. doi:10.1016/j.conb.2021.10.010

Dahmen D, Recanatesi S, Jia X, Ocker GK, Campagnola L, Seeman S, Jarsky T, Helias M, Shea-Brown E. 2023. Strong and localized recurrence controls dimensionality of neural activity across brain areas. *bioRxiv* 2020.11.02.365072. doi:10.1101/2020.11.02.365072

Denève S, Alemi A, Bourdoukan R. 2017. The Brain as an Efficient and Robust Adaptive Learner. *Neuron* **94**:969–977. doi:10.1016/j.neuron.2017.05.016

Denève S, Machens CK. 2016. Efficient codes and balanced networks. *Nat Neurosci* **19**:375–382. doi:10.1038/nn.4243

Denk W, Briggman KL, Helmstaedter M. 2012. Structural neurobiology: missing link to a mechanistic understanding of neural computation. *Nat Rev Neurosci* **13**:351–358. doi:10.1038/nrn3169

Festa D, Hennequin G, Lengyel M. 2018. Analog Memories in a Balanced Rate-Based Network of E-I Neurons 1–9.

Frank T, Mönig1 NR, Satou1 C, Friedrich SH Rainer W, Friedrich RW. 2019. Associative conditioning remaps odor representations and modifies inhibition in a higher olfactory brain area. *Nat Neurosci* 1–12. doi:10.1038/s41593-019-0495-z

Franks KM, Russo MJ, Sosulski DL, Mulligan AA, Siegelbaum SA, Axel R. 2011. Recurrent Circuitry Dynamically Shapes the Activation of Piriform Cortex. *Neuron* **72**:49–56. doi:10.1016/j.neuron.2011.08.020

Freeman WJ, Skarda CA. 1985. Spatial EEG patterns, non-linear dynamics and perception: the neo-sherringtonian view. *Brain Res Rev* **10**:147–175. doi:10.1016/0165-0173(85)90022-0

Friedrich RW, Laurent G. 2004. Dynamics of Olfactory Bulb Input and Output Activity During Odor Stimulation in Zebrafish. *Journal of Neurophysiology* **91**:2658–2669. doi:10.1152/jn.01143.2003

Friedrich RW, Laurent G. 2001. Dynamic Optimization of Odor Representations by Slow Temporal Patterning of Mitral Cell Activity. *Science* **291**:1–7.

Friedrich RW, Wanner A. 2021. Dense Circuit Reconstruction to Understand Neuronal Computation: Focus on Zebrafish. *Annual Review of Neuroscience* **44**:1–19. doi:10.1146/annurev-neuro-110220-013050

Froemke RC, Merzenich MM, Schreiner CE. 2007. A synaptic memory trace for cortical receptive field plasticity. *Nature* **450**:425–429. doi:10.1038/nature06289

Gallego JA, Perich MG, Miller LE, Solla SA. 2017. Neural Manifolds for the Control of Movement. *Neuron* **94**:978–984. doi:10.1016/j.neuron.2017.05.025

Ghojogh B, Crowley M. 2019. Linear and Quadratic Discriminant Analysis: Tutorial. *arXiv*. doi:10.48550/arxiv.1906.02590

Haberly LB. 2001. Parallel-distributed Processing in Olfactory Cortex: New Insights from Morphological and Physiological Analysis of Neuronal Circuitry. *Chem Senses* **26**:551–576. doi:10.1093/chemse/26.5.551

Hebb DO. 1949. The organization of behavior; a neuropsychological theory. Wiley.

Hennequin G, Agnes EJ, Vogels TP. 2017. Inhibitory Plasticity: Balance, Control, and Codependence. *Annu Rev Neurosci* **40**:557–579. doi:10.1146/annurev-neuro-072116-031005

Hennequin G, Vogels TP, Gerstner W. 2014. Optimal Control of Transient Dynamics in Balanced Networks Supports Generation of Complex Movements. *Neuron* **82**:1394–1406. doi:10.1016/j.neuron.2014.04.045

Hopfield JJ. 1982. Neural networks and physical systems with emergent collective computational abilities. *Proc Natl Acad Sci* 2554–2558.

Iurilli G, Datta SR. 2017. Population Coding in an Innately Relevant Olfactory Area. *Neuron* **93**:1180-1197.e7. doi:10.1016/j.neuron.2017.02.010

Jacobson GA, Rupprecht P, Friedrich RW. 2018. Experience-Dependent Plasticity of Odor Representations in the Telencephalon of Zebrafish. *Curr Biol* **28**:1-14.e3. doi:10.1016/j.cub.2017.11.007

Khona M, Fiete IR. 2022. Attractor and integrator networks in the brain. *Nat Rev Neurosci* **23**:744–766. doi:10.1038/s41583-022-00642-0

Kohonen T. 1984. Self-Organzation and Associative Memory.

Kornfeld J, Denk W. 2018. Progress and remaining challenges in high-throughput volume electron microscopy. *Curr Opin Neurobiol* **50**:261–267. doi:10.1016/j.conb.2018.04.030

Lagzi F, Fairhall A. 2022. Tuned inhibitory firing rate and connection weights as emergent network properties. *bioRxiv* 2022.04.12.488114. doi:10.1101/2022.04.12.488114

Langdon C, Genkin M, Engel TA. 2023. A unifying perspective on neural manifolds and circuits for cognition. *Nat Rev Neurosci* **24**:363–377. doi:10.1038/s41583-023-00693-x

Litwin-Kumar A, Doiron B. 2014. Formation and maintenance of neuronal assemblies through synaptic plasticity. *Nature Communications* **5**:1–12. doi:10.1038/ncomms6319

Litwin-Kumar A, Doiron B. 2012. Slow dynamics and high variability in balanced cortical networks with clustered connections. *Nat Neurosci* **15**:1498–1505. doi:10.1038/nn.3220

Mackwood O, Naumann LB, Sprekeler H. 2021. Learning excitatory-inhibitory neuronal assemblies in recurrent networks. *Elife* **10**:e59715. doi:10.7554/elife.59715

Marr D. 1969. A theory of cerebellar cortex. *J Physiol* **202**:437–470.

Meissner-Bernard C, Dembitskaya Y, Venance L, Fleischmann A. 2018. Encoding of Odor Fear Memories in the Mouse Olfactory Cortex. *Current Biology* **29**:1–19. doi:10.1016/j.cub.2018.12.003

Miehl C, Onasch S, Festa D, Gjorgjieva J. 2022. Formation and computational implications of assemblies in neural circuits. *J Physiology*. doi:10.1113/jp282750

Miura K, Mainen ZF, Uchida N. 2012. Odor Representations in Olfactory Cortex: Distributed Rate Coding and Decorrelated Population Activity. *Neuron* **74**:1087–1098. doi:10.1016/j.neuron.2012.04.021

Mueller T, Dong Z, Berberoglu MA, Guo S. 2011. The dorsal pallium in zebrafish, Danio rerio (Cyprinidae, Teleostei). *Brain Res* **1381**:95–105. doi:10.1016/j.brainres.2010.12.089

Murphy BK, Miller KD. 2009. Balanced Amplification: A New Mechanism of Selective Amplification of Neural Activity Patterns. *Neuron* **61**:635–648. doi:10.1016/j.neuron.2009.02.005

Okun M, Lampl I. 2008. Instantaneous correlation of excitation and inhibition during ongoing and sensory-evoked activities. *Nat Neurosci* **11**:535–537. doi:10.1038/nn.2105

Pashkovski SL, Iurilli G, Brann D, Chicharro D, Drummey K, Franks K, Panzeri S, Datta SR. 2020. Structure and flexibility in cortical representations of odour space. *Nature* 1–28. doi:10.1038/s41586-020-2451-1

Pehlevan C, Sompolinsky H. 2014. Selectivity and Sparseness in Randomly Connected Balanced Networks 1–15.

Recanatesi S, Ocker GK, Buice MA, Shea-Brown E. 2019. Dimensionality in recurrent spiking networks: Global trends in activity and local origins in connectivity. *PLoS Comput Biol* **15**:e1006446. doi:10.1371/journal.pcbi.1006446

Renart A, Rocha J de la, Bartho P, Hollender L, Parga N, Reyes A, Harris KD. 2010. The Asynchronous State in Cortical Circuits. *Science* **327**:587–590. doi:10.1126/science.1179850

Rost T, Deger M, Nawrot MP. 2018. Winnerless competition in clustered balanced networks: inhibitory assemblies do the trick. *Biol Cybern* **112**:81–98. doi:10.1007/s00422-017-0737-7

Roudi Y, Latham PE. 2007. A Balanced Memory Network. *PLoS Comput Biol* **3**:e141-22. doi:10.1371/journal.pcbi.0030141

Rupprecht P, Carta S, Hoffmann A, Echizen M, Blot A, Kwan AC, Dan Y, Hofer SB, Kitamura K, Helmchen F, Friedrich RW. 2021. A database and deep learning toolbox for noise-optimized, generalized spike inference from calcium imaging. *Nat Neurosci* **24**:1324–1337. doi:10.1038/s41593-021-00895-5

Rupprecht P, Friedrich RW. 2018. Precise Synaptic Balance in the Zebrafish Homolog of Olfactory Cortex. *Neuron* **100**:1–29. doi:10.1016/j.neuron.2018.09.013

Ryan TJ, Roy DS, Pignatelli M, Arons A, Tonegawa S. 2015. Engram Cells Retain Memory Under Retrograde Amnesia. *Science* 1007–1013.

Sacco T, Sacchetti B. 2010. Role of Secondary Sensory Cortices in Emotional Memory Storage and Retrieval in Rats. *Science* **329**:649–656. doi:10.1126/science.1183165

Sadeh S, Clopath C. 2021. Excitatory-inhibitory balance modulates the formation and dynamics of neuronal assemblies in cortical networks. *Sci Adv* **7**:eabg8411. doi:10.1126/sciadv.abg8411

Sadeh S, Clopath C. 2020a. Patterned perturbation of inhibition can reveal the dynamical structure of neural processing. *eLife* **9**:226–29. doi:10.7554/elife.52757

Sadeh S, Clopath C. 2020b. Inhibitory stabilization and cortical computation. *Nat Rev Neurosci* 1–17. doi:10.1038/s41583-020-00390-z

Schaffer ES, Stettler DD, Kato D, Choi GB, Axel R, Abbott LF. 2018. Odor Perception on the Two Sides of the Brain: Consistency Despite Randomness. *Neuron* **98**:736-742.e3. doi:10.1016/j.neuron.2018.04.004

Schoonover CE, Ohashi SN, Axel R, Fink AJP. 2021. Representational drift in primary olfactory cortex. *Nature* 1–34. doi:10.1038/s41586-021-03628-7

Schulz A, Miehl C, II MJB, Gjorgjieva J. 2021. The generation of cortical novelty responses through inhibitory plasticity. *eLife* **10:e65309**:1–28. doi:10.1101/2020.11.30.403840

Shadlen MN, Newsome WT. 1994. Noise, neural codes and cortical organization. *Curr Opin Neurobiol* **4**:569–579. doi:10.1016/0959-4388(94)90059-0

Stettler DD, Axel R. 2009. Representations of Odor in the Piriform Cortex. *Neuron* **63**:854–864. doi:10.1016/j.neuron.2009.09.005

Tabor R, Friedrich RW. 2008. Pharmacological Analysis of Ionotropic Glutamate Receptor Function in Neuronal Circuits of the Zebrafish Olfactory Bulb. *PLoS ONE* **3**:e1416. doi:10.1371/journal.pone.0001416

Tabor R, Yaksi E, Weislogel J-M, Friedrich RW. 2004. Processing of Odor Mixtures in the Zebrafish Olfactory Bulb. *The Journal of Neuroscience* **24(29)**:6611– 6620. doi:10.1523/jneurosci.1834-04.2004

Van Vreeswijk C, Sompolinsky H. 1996. Chaos in Neuronal Networks with Balanced Excitatory and Inhibitory Activity. *Science* **274**:1724–1726. doi:10.1126/science.274.5293.1724

Vogels TP, Sprekeler H, Zenke F, Clopath C, Gerstner W. 2011. Inhibitory Plasticity Balances Excitation and Inhibition in Sensory Pathways and Memory Networks. *Science* **334**:1–7. doi:10.1126/science.1212991

Wanner AA, Friedrich RW. 2020. Whitening of odor representations by the wiring diagram of the olfactory bulb. *Nat Neurosci* 1–24. doi:10.1038/s41593-019-0576-z

Wehr M, Zador AM. 2003. Balanced inhibition underlies tuning and sharpens spike timing in auditory cortex. *Nature* **426**:1–5.

Wiechert MT, Judkewitz B, Riecke H, Friedrich RW. 2010. Mechanisms of pattern decorrelation by recurrent neuronal circuits. *Nat Neurosci* **13**:1003–1010. doi:10.1038/nn.2591

Wilson DA, Sullivan RM. 2011. Cortical Processing of Odor Objects. *Neuron* **72**:506–519. doi:10.1016/j.neuron.2011.10.027

Wu YK, Zenke F. 2021. Nonlinear transient amplification in recurrent neural networks with short-term plasticity. *eLife* 2021.06.09.447718. doi:10.1101/2021.06.09.447718

Yaksi E, Paul F von S, Niessing J, Bundschuh ST, Friedrich RW. 2009. Transformation of odor representations in target areas of the olfactory bulb. *Nat Neurosci* **12**:474–482. doi:10.1038/nn.2288

Zenke F, Agnes EJ, Gerstner W. 2015. Diverse synaptic plasticity mechanisms orchestrated to form and retrieve memories in spiking neural networks. *Nature Communications* **6**:1–13. doi:10.1038/ncomms7922

Zou M. 2014. Connectivity, Plasticity, and Function of Neuronal Circuits in the Zebrafish Olfactory Forebrain.

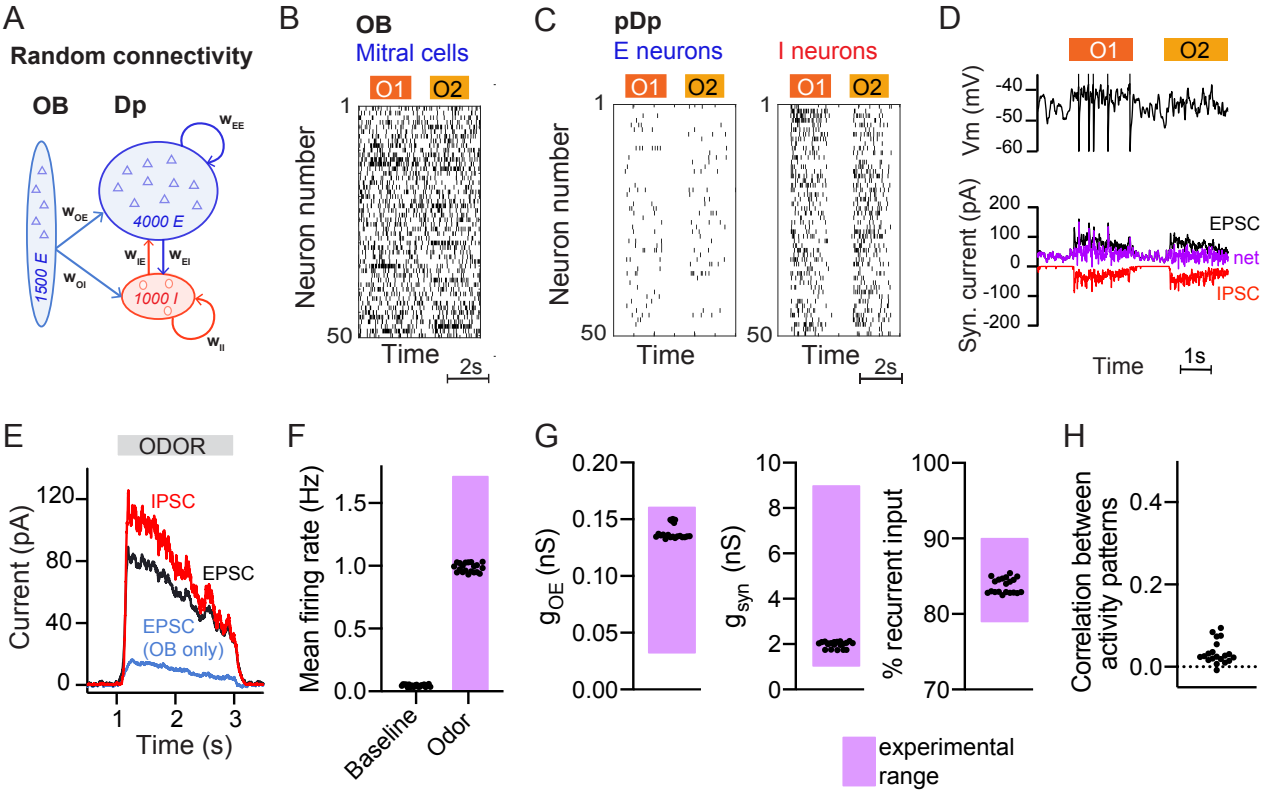


Figure 1. Spiking network model of pDp. A. Schematic of pDp_{sim}. OB, Olfactory bulb; E, excitatory; I, inhibitory neurons. **B.** Spike raster of a random subset of 50 mitral cells in the OB representing 2 odors. During odor stimulation, firing rates of 10% of mitral cells were increased and firing rates of 5% of mitral cells were decreased (baseline rate, 6 Hz). **C.** Spike raster of random subsets of 50 E and I neurons in response to 2 odors. **D.** Representative membrane potential trace (top) and excitatory (EPSC, black) and inhibitory (IPSC, red) currents (bottom) in one excitatory neuron in response to 2 odors. Purple trace shows net current (EPSC + IPSC). **E.** Odor-evoked inhibitory (red) and excitatory (black and blue) currents as measured in a hypothetical voltage clamp experiment (conductance multiplied by 70 mV, the absolute difference between holding potential and reversal potential; Rupprecht et al., 2018). Representative example of 1 network, averaged across neurons and odors. **F-H**. Measured values of the observables used to match pDp_{sim} to experimental data. Each dot represents one network (average over 10 odors); n = 20 networks. Pink shading shows the experimentally observed range of values. **F.** Baseline and odor-evoked population firing rate. **G.** Left: g_{OE} is the synaptic conductance in E neurons contributed by afferents from the OB during odor stimulation. Middle: g_{syn} is the total odor-evoked synaptic conductance. Right: % recurrent input quantifies the percentage of E input contributed by recurrent connections during odor stimulation. **H.** Correlation coefficient between odor-evoked activity patterns in Dp. The dotted line indicates the mean correlation between odor patterns in the OB.

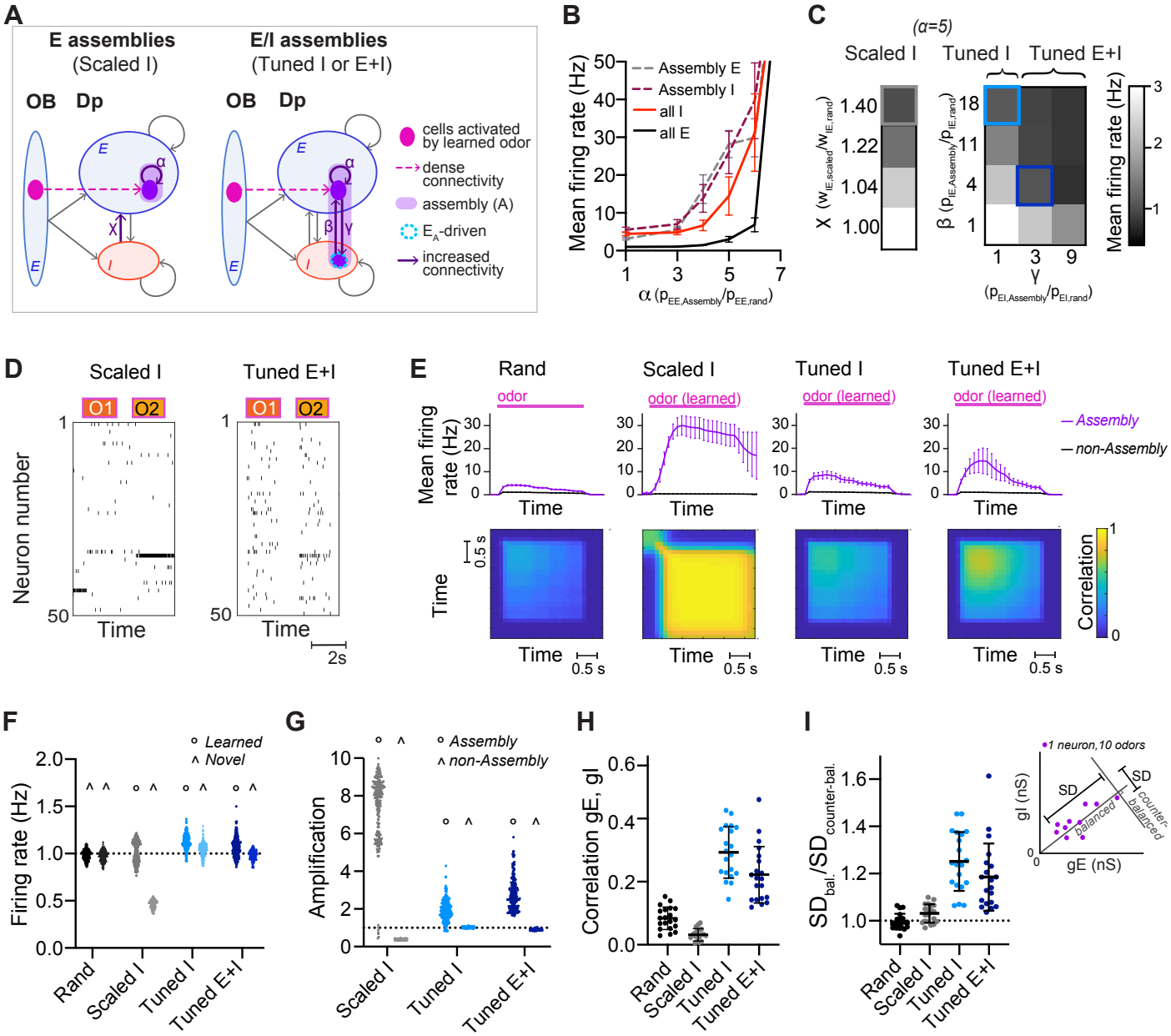


Figure 2. Neuronal assemblies (memories). A. Schematic of assemblies. Each assembly contains the 100 E neurons that are most densely connected to the mitral cells activated by a given odor. Connection probability between these E neurons is increased by a factor α. In Scaled I networks, weights of all I-to-E synapses are increased by a factor χ. In Tuned networks, the 25 I neurons that are most densely connected to the 100 E neurons are included in each assembly. In Tuned I networks, the probability of I-to-E connections within the assembly is increased by a factor β. In Tuned I+E networks, probabilities of I-to-E and E-to-I connectivity within the assembly are increased by factors β and γ, respectively. n = 20 networks with 15 assemblies each were simulated for each group. **B.** Firing rates averaged over all E or I neurons (full lines) and over all assembly neurons (dashed lines) as a function of α (mean ± SD. across 20 networks). **C.** Mean E neurons firing rates of Scaled (left) and Tuned (right) networks in response to learned odors as a function of connection strength and probability, respectively. Squares depict parameters used in following figures unless stated otherwise. **D.** Spike raster plots showing responses of 50 E neurons to two odors in a Scaled I and the corresponding Tuned E+I network (same neurons and odors in the corresponding rand network are shown in Fig. 1C). **E.** Top: Mean firing rates in response to learned odors as a function of time, averaged across assembly or non-assembly neurons. Bottom: Correlation between activity patterns evoked by different trials of the same learned odor as a function of time. The pink bar indicates odor presentation. **F.** Mean firing rate in response to learned odors or novel odors. Each data point represents one network-odor pair (n=20 networks, 10 odors). **G.** Amplification within and outside (non-A.) assemblies, calculated as the ratio between mean firing rates in response to learned odors averaged across the same populations of neurons in a given structured network (Scaled I, Tuned I, or Tuned E+I) and the corresponding rand network. **H.** Quantification of co-tuning by the correlation between time-averaged E and I conductances in response to different odors, average across neurons (n = 20 networks). **I.** Quantification of co-tuning by the ratio of dispersion of joint conductances along balanced and counter-balanced axes (Methods). Each data point corresponds to one network (n = 20). Mean +/- SD.

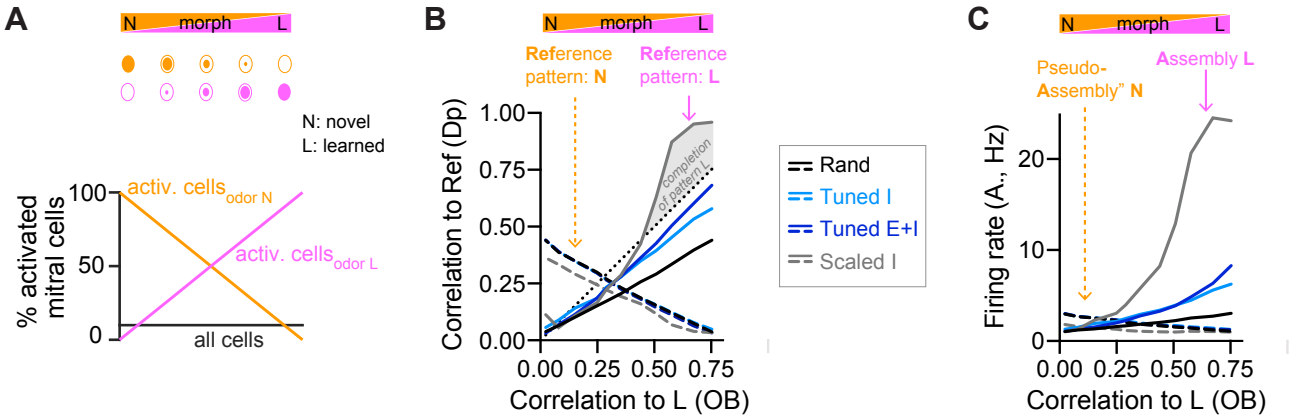


Figure 3. Changes of output activity to gradual modifications of inputs. A. Morphing of a novel odor N into a learned odor L. Intermediate mixtures were generated by gradually decreasing/increasing the fractions of active mitral cells defining odors N/L. **B.** Pearson correlation between activity patterns evoked by the learned (full line) or novel (dotted line) odor and the intermediate odors in pDp as a function of the corresponding correlations in the OB. **C.** Firing rates in response to intermediate odors averaged across assembly neurons (learned odor, full line) or pseudo-assembly neurons (novel odor, dotted line) as a function of correlations between the OB activity patterns representing the intermediate odors and the OB activity pattern representing the learned odor. B, C: averages over 8 networks.

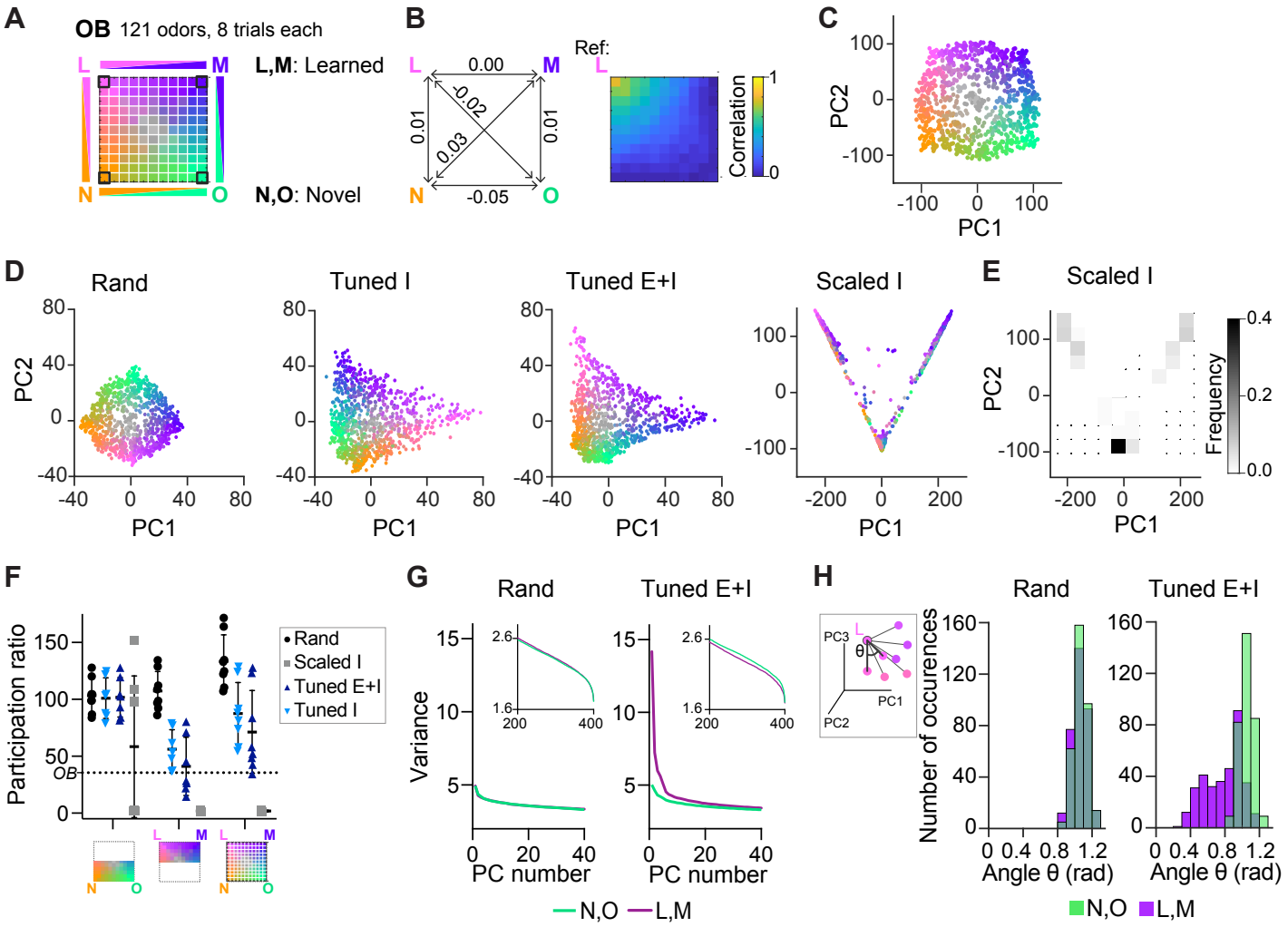


Figure 4. Geometry of odor representations in pDp_{sim}. A. Odor subspace delineated by 2 learned (L, M) and 2 novel (N, O) pure odors at the vertices of a square. Pixels within the square represent odor mixtures. **B.** Left: Pearson correlations between OB activity patterns defining pure odors. Right: Correlation between one pure odor (L; top left vertex) and all other odors. The odor from one vertex gradually morphs into the odor from another vertex. **C.** Projection of activity patterns in the OB onto the first 2 principal components (PCs). Colors represent patterns in the odor subspace shown in A. **D.** Projection of activity patterns in pDp in response to the odor subspace onto the first 2 PCs. Representative examples of different pDp networks. **E.** Density plot showing distribution of data points and demonstrating clustering at distinct locations in PC space for Scaled I networks. **F.** Quantification of dimensionality of neural activity by the participation ratio: activity evoked by novel odors and related mixtures (left), activity evoked by learned odors and related mixtures (center), and activity evoked by all stimuli (right). Each data point represents one network; dotted line represents the participation ratio of OB activity. **G.** Variance along the first 40 PCs extracted from activity patterns in Rand and Tuned E+I networks. Insets: variance along PCs 200 - 400. **H.** Angles between edges connecting a pure odor response and related versions thereof (see inset). The analysis was performed using the first 400 PCs, which explained >75% of the variance in all networks. n = 21 angles per pure odor in each of 8 networks (Methods). Similar results were obtained in the full-dimensional space.

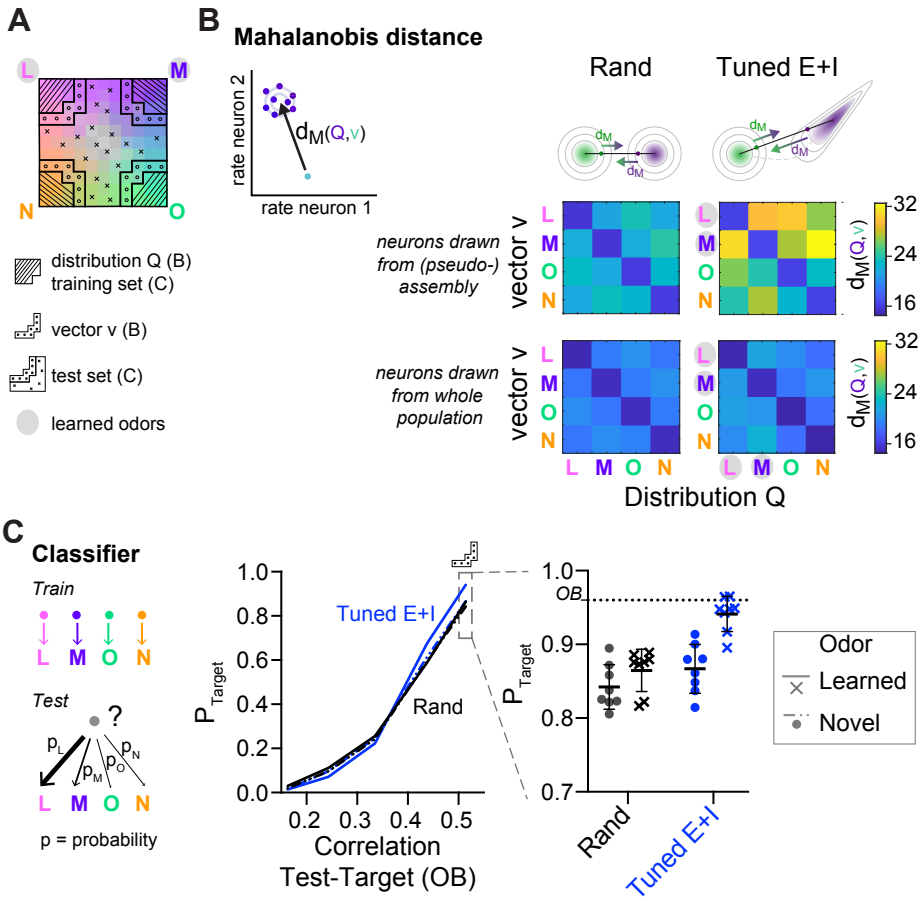


Figure 5. Distance relationships and classification of odor representations. A. Activity patterns used as class distributions and vectors (B) or training and test sets (C). Same odor subspace as in Figure 4. **B.** Left: schematic illustration of Mahalanobis distance dM. Right: dM between one activity vector (v) and reference classes (Q) in rand and Tuned E+I networks. dM was computed based on activity across subsets of 80 E neurons drawn from the four (pseudo) assemblies with equal probability (top) or from the whole population (bottom). Note that dM between patterns related to a learned odor and non-matching reference classes was higher in Tuned E+I networks, particularly when E neurons were drawn from assemblies. **C.** Pattern classification probability quantified by QDA. P_{Target} quantifies the probability that an activity pattern from the test set (odor mixtures, see A) is assigned to a target class from the training set (pure or closely related odor; see A). Left: classification probability as a function of the similarity (Pearson correlation) between the test and target odors in the OB (input patterns). Note enhanced classification probability for patterns evoked by odors similar to learned odors in Tuned E+I networks. Right: Classification probability for patterns similar to the training set (see A).

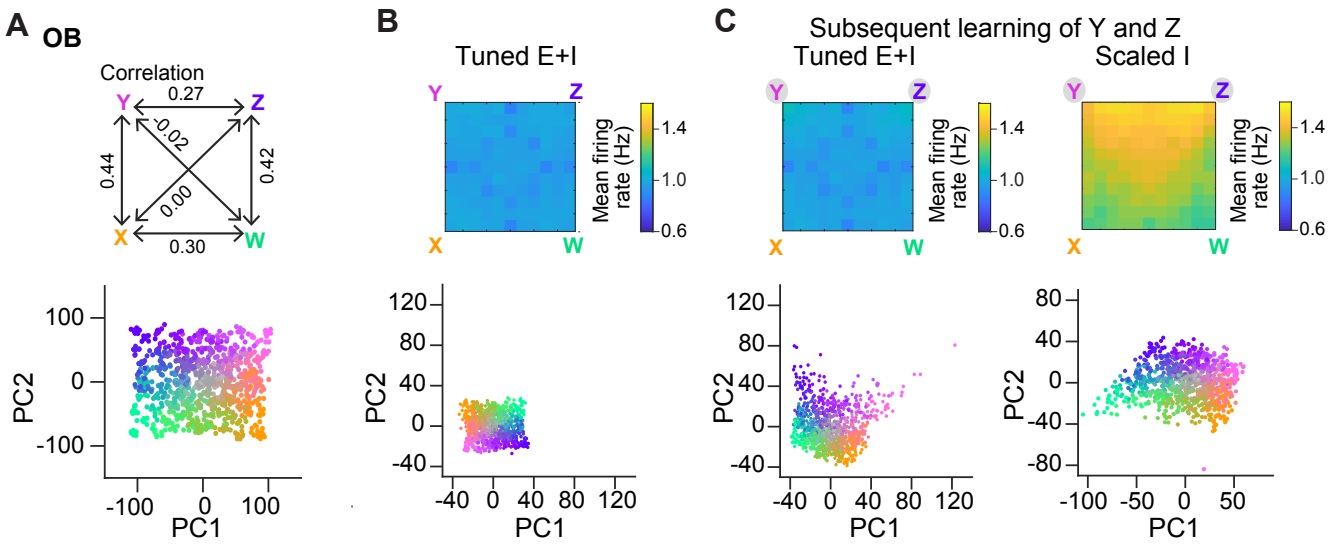


Figure 6. Representation of correlated patterns and resilience against additional memories. A. Subspace delineated by four positively correlated odors. Top: Correlations between pure odors. Bottom: Projection of OB activity patterns onto the first two PCs. **B.** Firing rates (top) and PC projection of output activity of a Tuned E+I network with 15 E/I assemblies that were not aligned to any of the four pure odors of the subspace. **C.** Firing rates (top) and PC projection of output activity after creation of two additional assemblies representing two of the pure odors (Y and Z). Left: Tuned E+I network. Right: Scaled I network. Note that in the Scaled I network, but not in the Tuned E+I network, firing rates evoked by newly learned odors were increased and patterns evoked by these odors were not well separated in PC space.

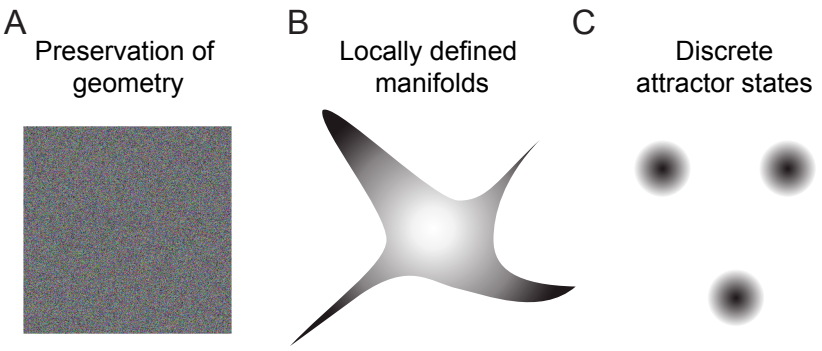
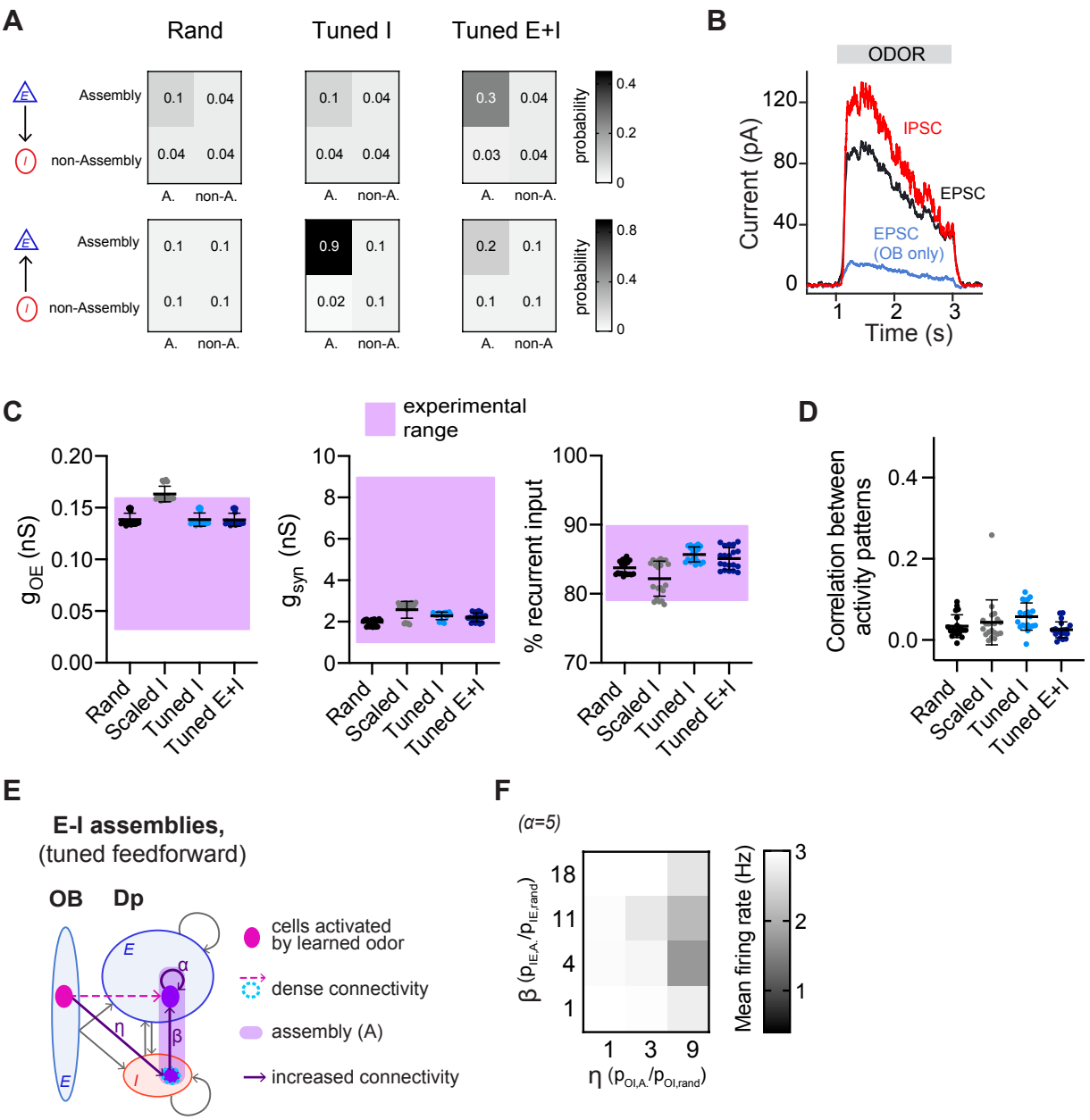
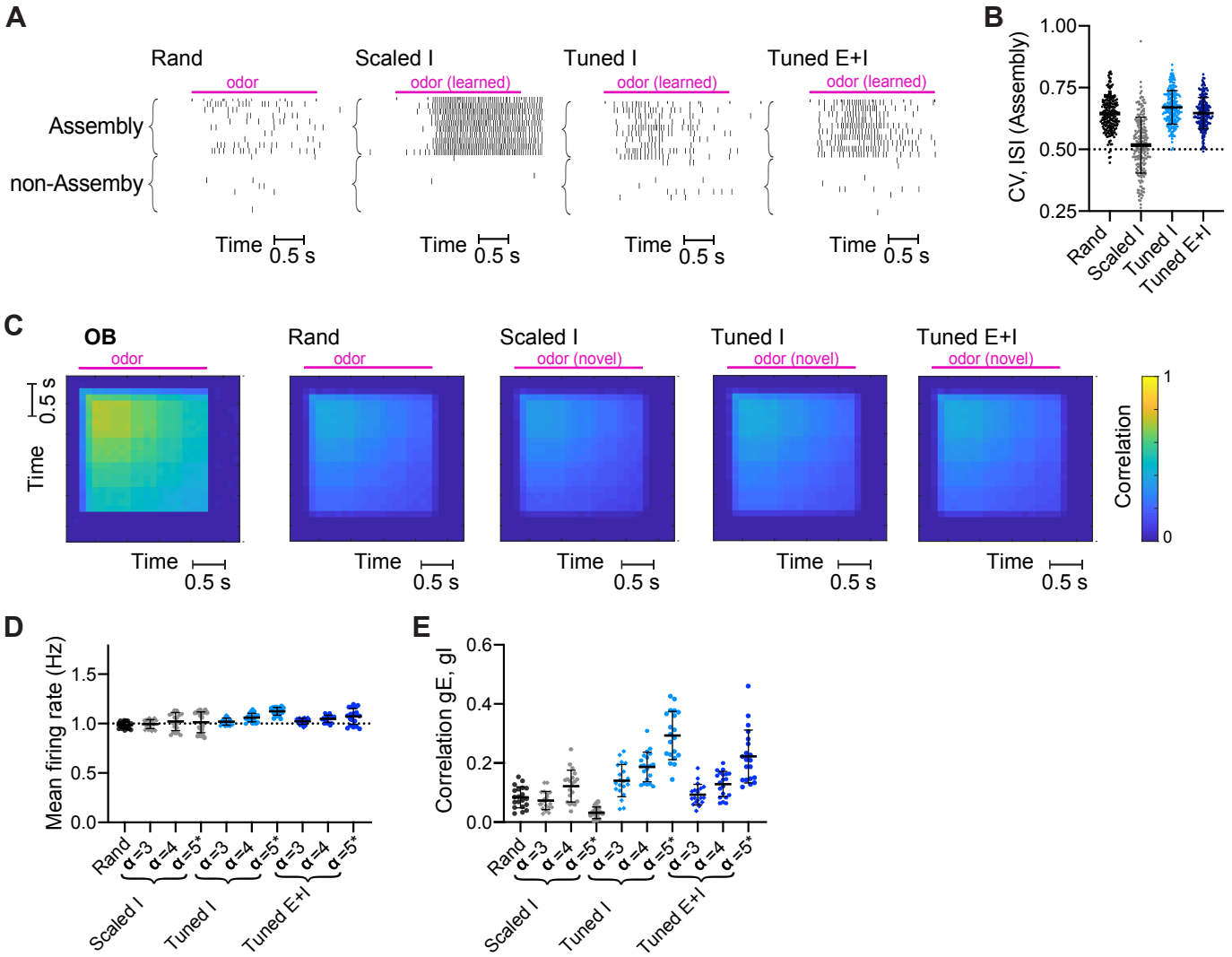


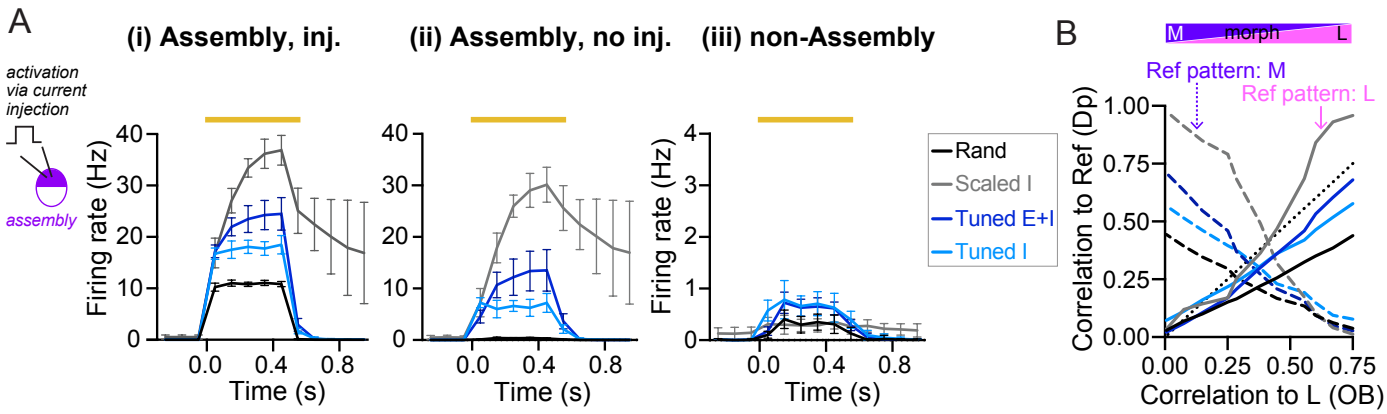
Figure 7. Schematic of geometric transformations. A. Randomly connected networks tend to preserve the geometry of coding space. Such networks can support neuronal computations, e.g., by projecting activity patterns in a higher-dimensional coding space for pattern classification. **B.** We found that balanced networks with E/I assemblies transform the geometry of representations by locally restricting activity onto manifolds. These networks stored information about learned inputs while preserving continuity of the coding space. Such a geometry may support fast classification, continual learning and cognitive computations. Note that the true manifold geometry cannot be visualized appropriately in 2D because activity was "focused" in different subsets of dimensions at different locations of coding space. As a consequence, the dimensionality of activity remained substantial. **C.** Neuronal assemblies without precise balance established discrete attractor states, as observed in memory networks that store information as discrete items. Networks establishing locally defined activity manifolds (B) may thus be considered as intermediates between networks generating continuous representations without memories (A) and classical memory networks with discrete attractor dynamics (C).



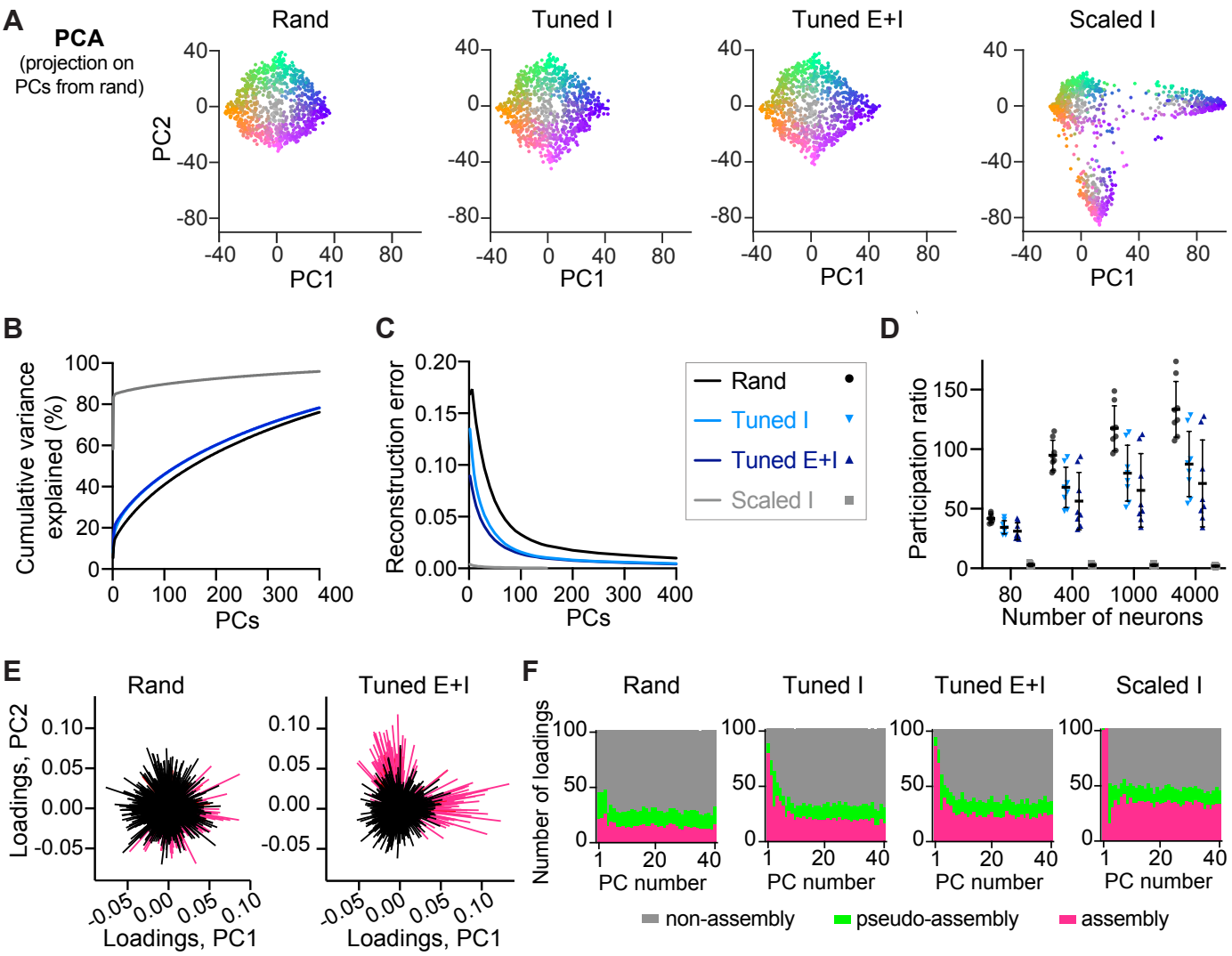
Supplementary Figure 2I. Structured networks reproduce key features of Dp. A. Connection probability between classes of E and I neurons in a rand, a Tuned I, and a Tuned E+I network. **B.** EPSCs and IPSCs in a Tuned I network as observed in a hypothetical voltage clamp recording, averaged across neurons and odors. An equivalent plot for a rand network is shown in Figure 1E. **C,D.** Values of observables for rand networks (same as Figure 1G-H) and different structured networks (Scaled I, Tuned I, Tuned E+I). **E.** Network with increased connectivity between the E assembly neurons (α=5) and the 25 I neurons that are most densely connected to the mitral cells activated by a given odor. **F.** Mean firing rate of the network in E in response to learned odors as a function of connection probability. Selecting I neurons based on their afferent connectivity could not stabilize activity efficiently.



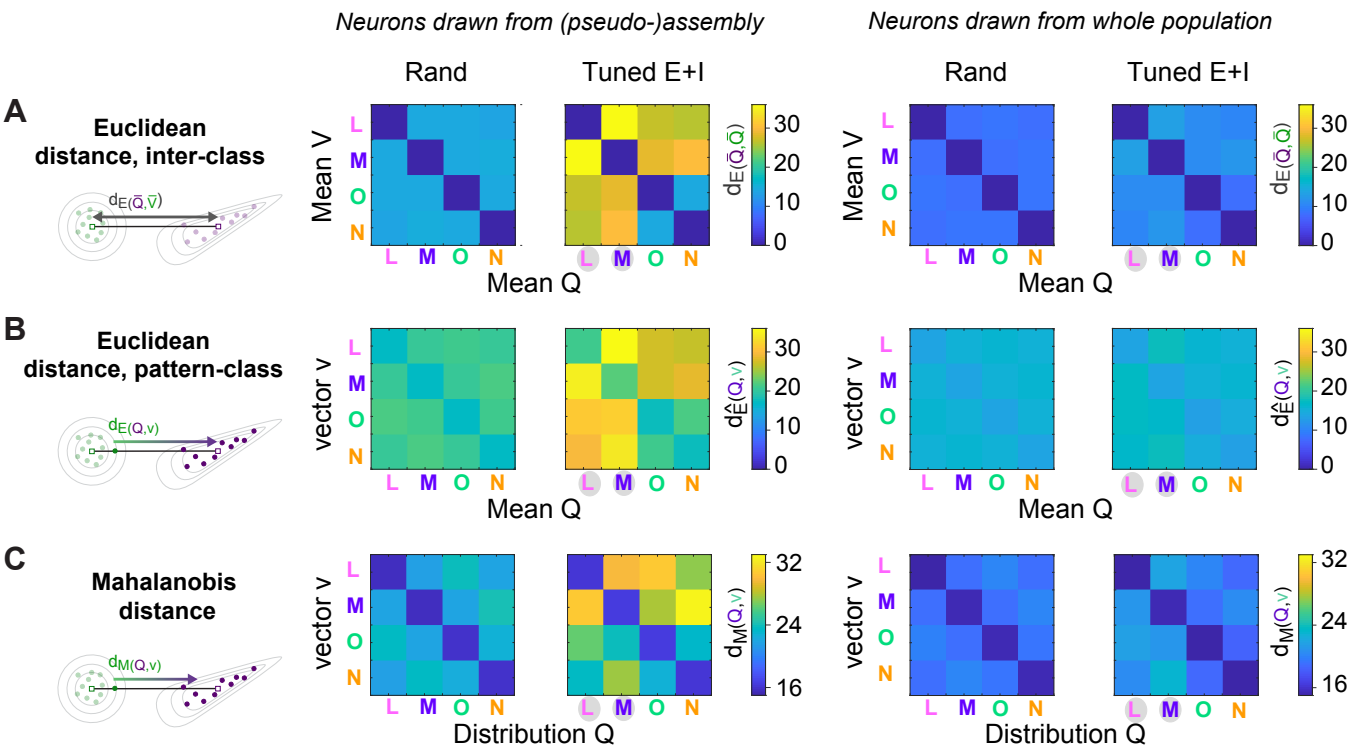
Supplementary Figure 2II. Structured networks: additional results. A. Raster plots showing responses of assembly (A) or non-assembly neurons to a learned odor. **B.** Coefficients of variation of the inter-spike interval (ISI) in assembly neurons. **C.** Correlation between activity patterns across E neurons evoked by the same novel odor in different trials as a function of time. Pink bar indicates odor presentation. Note that correlations in response to novel odors are similar across networks and different from responses to learned odors in Scaled and Tuned networks (Fig. 2E). **D.** Population firing rate in response to learned odors for networks with different probability of E-E connectivity within assemblies (α). **E.** Co-tuning (correlation between E and I currents in individual neurons) as a function of α.



Supplementary Figure 3. Pattern completion: additional results. A. Artificial reactivation of E assemblies. During 6 Hz baseline activity of the olfactory bulb, a subset of the assembly neurons was artificially reactivated by current injection (500 ms, 28 nA). Mean firing rates were quantified in the injected assembly neurons (i), in the remaining, non-injected assembly or pseudo-assembly neurons (ii), and in the non-assembly neurons (iii) as a function of time. The orange bar indicates duration of current injection. **B.** Correlation between activity patterns across E neurons (output patterns) evoked by a series of input patterns representing a morph of one learned odor into another learned odor. Correlations between output patterns are plotted as a function of the correlation between the corresponding OB patterns.



Supplementary Figure 4. Transformations and dimensionality of activity patterns: additional results. A. Projection of activity patterns representing the odor subspace (Fig. 4A) onto the first 2 PCs of the corresponding rand networks (representative examples of one network each). **B.** Scree plot for the PCA results shown in Figure 4D. **C.** Error in the reconstruction of odor-evoked activity patterns as a function of the number of PCs. Euclidean distances between all pairs of activity patterns were calculated in the full-dimensional state space (Df) and in reduced-dimensional embedding spaces (Dl). The reconstruction error was defined as 1 – (correlation between Df and Dl). **D.** Participation ratio of the neural activity sampled from different numbers of neurons (50 iterations). **E.** Loadings of neurons on the first two PCs of a rand and a Tuned E+I network (Figure 4D). Each line represents one neuron. Neurons that are part of the assemblies representing the two learned odors are color-coded in magenta. **F.** For each network and PC, the 100 E neurons with the highest absolute loadings were selected and grouped into 3 categories: neurons part of the two assemblies representing the learned odors, neurons part of the two “pseudo-assemblies” representing the two novel odors, and the remaining neurons (non-assembly).



Supplementary Figure. 5. Further analyses of pattern distances. Differences in d_M between *rand* and *Tuned* networks may involve differences in the distance between class centers and/or differences in intra-class variability. To dissect the contributions of these effects we compared three distance measures:
(1) The Euclidean distance between class centers:

$$d_E = \sqrt{(\mu_V - \mu_Q)^T(\mu_V - \mu_Q)}$$

(2) The mean Euclidean distance between patterns of one class and the center of another class ($d_{\hat{E}}$):

$$d_{\hat{E}} = \frac{1}{n}\sum_{i=1..n}\sqrt{(v_i - \mu_Q)^T(v_i - \mu_Q)}$$

(3) The Mahalanobis distance d_M:

$$d_M = \frac{1}{n}\sum_{i=1..n}\sqrt{(v_i - \mu_Q)^T S^{-1}(v_i - \mu_Q)}$$

Here, μ_V and μ_Q are the centers (average patterns) of classes V and Q, respectively; v is a vector from V = [$v_1, v_2, \ldots, v_n$], and S^{-1} is the inverse of the covariance matrix of the neuronal population within reference class Q. Note that $d_{\hat{E}}$ corresponds to d_M without normalization by variability (covariance).
A. Analysis of d_E. Left: d_E in *rand* and *Tuned E+I* networks based on 80 E neurons drawn from (pseudo-) assemblies. Right: same based on 80 E neurons drawn from the entire population. Note that d_E between learned and other odors was increased in *Tuned E+I* networks as compared to *rand* networks, particularly when neurons were drawn from assemblies. **B**. Equivalent plots for $d_{\hat{E}}$. Note that distances were increased nearly symmetrically, similar to d_E. **C**. Equivalent plots for d_M (same plots as in Fig. 5B). Note that d_M was increased asymmetrically. These observations show that the changes in d_M relative to *rand* networks involved an increase in the distance between class centers (d_E) and a non-isotropic change in intra-class variability (comparison between $d_{\hat{E}}$ and d_M). These effects were prominent when E neurons were drawn from assemblies. An important contribution to the increase in d_M in the direction from learned odors to reference classes representing novel odors was made by the increased distance between class centers. In the other direction, d_M was smaller, implying that variability in the reference class was higher. Nonetheless, variability in the relevant direction did not fully counteract the increased distance between class centers in *Tuned E+I* networks. As a consequence, d_M was still increased slightly relative to the corresponding *rand* networks. Most of these effects were still observed, albeit weakly, when E neurons were drawn from the whole population.