Title:

Structured connectivity exploits NMDA non-linearitites to induce dimensionally diverse responses in a PFC circuit

Authors:

Stefanou S. Stefanos[1,2], Athanasia Papoutsi[1], Panayiota Poirazi[1,3,*]

Affiliations:

[1]*Institute of Molecular Biology and Biotechnology (IMBB), Foundation for Research and Technology-Hellas (FORTH), Heraklion, Crete, 70013, Greece*
[2] *Department of Biology, University of Crete, Heraklion, Crete, 70013, Greece*

Author list footnotes:
[3]Lead Contact
*Correspondence: *poirazi@imbb.forth.gr*

Summary:

Prefrontal Cortex (PFC) exerts control on action selection and mediates behavioral flexibility that is imperative during working memory, when integration and retention of stimuli takes place. During a working memory task, neurons of the PFC rapidly modify their response properties in a context-dependent manner. Could this flexible tuning be mediated by temporary connectivity states, supported by single-cell mechanisms? Are these states bound by the underlying network connectivity? Indeed, pyramidal neurons of the PFC are characterized by prominent non-linear dendritic responses and participate in hyper-reciprocal networks. This work probes these questions using a detailed, PFC model network equipped with nonlinear synaptic mechanisms. The model reproduces a wide range of experimental findings with respect to the response properties of individual cells and the dynamic population responses. We propose that PFC neurons dynamically form functional assemblies by transiently recruiting non-linear dendritic phenomena, shaping a robust yet flexible population response.

Keywords: prefrontal, working memory, dendritic integration, NMDA, behavioral flexibility, assemblies, dimensionality

Introduction

Working memory (WM) is the ability to maintain behaviorally relevant stimuli for a short period of time and to integrate them towards a choice/action. Seminal WM studies observed that single-cell neuronal activity increases during the presentation of a to-be-maintained item (stimulus epoch) and persists throughout the blank 'post-stimulus' interval (post-stimulus epoch). This activity bridges the temporal gap between the item presentation and the subsequent contingent response (Funahashi, Bruce, & Goldman-Rakic, 1989; Fuster & Alexander, 1971), constitutes a basic component of decision making and is involved in the top-down modulation of behavior (Van Kerkoerle, Self, & Roelfsema, 2017). The above findings form the backbone of the so-called fixed selectivity framework that has dominated the WM field for many decades. According to this framework, selective tuning of neurons or neuronal populations to the to-be-remembered information can be explained by the formulation of stable points, based on the fine-tuning of the respective network connections, whereby individual neurons fire consistently to specific stimuli, both during the stimulus and the post-stimulus epoch (Renart, Song, & Wang, 2003; Wimmer, Nykamp, Constantinidis, & Compte, 2014).

Recent electrophysiological findings challenge this framework, noting that the dynamics underlying WM in the PFC are more diverse than previously thought. First, individual neurons display time-varying discharge rates (Bernacchia, Seo, Lee, & Wang, 2011; Brody, Hernández, Zainos, & Romo, 2003; Jun et al., 2010; Shafi et al., 2007), leading to the hypothesis that the tuning properties of neurons change throughout the task duration. Second, individual PFC neurons can be selective for a broad range of task variables, exhibiting thus a 'mixed selectivity' (Rigotti et al., 2013). Third, at the population level, the stimulus-induced network activity decays after the

stimulus offset and emerges later with a different population activity pattern (Stokes et al., 2013) yet containing a stable representation of the to-be-remembered item (Murray et al., 2017).

Single-neuron dynamic behavior and mixed selectivity has thus far been implemented using recurrent neural networks with random connectivity and non-biological plausible training of either the weights of the readout neuron or of the whole network (Barak, Rigotti, & Fusi, 2013; Barak, Sussillo, Romo, Tsodyks, & Abbott, 2013; Mante, Sussillo, Shenoy, & Newsome, 2013), thus constraining their biological- relevant predictions. In contrast to the training required by these approaches, it has been shown that PFC circuits typically utilize short-term dynamic mechanisms that renders them flexible (Bouchacourt & Buschman, 2019; Fujisawa, Amarasingham, Harrison, & Buzsáki, 2008). Finally, the representational shift at the population level from the sensory towards a memory state has thus far been attributed to either the dynamic behavior of a randomly connected recurrent neuron network (Bouchacourt & Buschman, 2019) or to short-term facilitation mechanisms (Barak, Tsodyks, & Romo, 2010). Both approaches depend on the implementation of distinct populations of sensory and memory neurons. Yet, selection and integration are implemented by the same PFC circuit, with both representations being present at the single neuron level (Mante et al., 2013). In this case, the population activity shift indicates integration of an incoming stimulus with its internal state that depends on short-term dynamics (Buonomano & Maass, 2009).

At these short-term time scales, it remains an open question how a recurrently connected network utilizes a rapid mechanism that renders it able a) to perform both selection and integration b) to be highly dynamic at the single neuron level yet c) stable and separable at the population level. Interestingly, theoretically work suggests that the highly non-random connectivity of pyramidal neurons, characterized by distance-dependent formation of neuronal clusters and over-represented

structural ‘motifs’ (structured connectivity) (Perin, Berger, & Markram, 2011; Song, Sjöström, Reigl, Nelson, & Chklovskii, 2005), is optimal for storing a large number of attractor states in a robust fashion (Brunel, 2016). In the PFC in particular, layer 5 (L5) pyramidal neurons have almost double the reciprocal connections of other cortical regions, forming hyper-reciprocally connected circuits (Morishima & Kawaguchi, 2006; Morishima, Morita, Kubota, & Kawaguchi, 2011; Y. Wang et al., 2006). This hyper-reciprocity could be a major player in the formation of multiple, transient, assemblies in the PFC. In addition to connectivity features, the NMDA receptor has been proposed to be a key mechanism for the formation of stable attractor dynamics (Durstewitz, Seamans, & Sejnowski, 2000; X.-J. Wang, 1999). The non-linearity and slow kinetics of the NMDA receptor makes it an alternative candidate to the short-term plasticity mechanisms, that can contribute to the activity-dependent selectivity and storage of the sensory input. In support, theoretical work proposed that NMDA-dependent dendritic non-linearities underlie the expression of memory states in small cell assemblies (Papoutsi, Kastellakis, & Poirazi, 2017; Papoutsi, Sidiropoulou, & Poirazi, 2014).

In this work we propose that structured connectivity coupled to NMDA nonlinearities enable PFC neurons to rapidly modify their response properties so as to form flexible representations of different inputs. To test this proposition, we use detailed biophysical models of PFC neuronal networks that are heavily constrained and show that they reproduce multiple types of experimental data. We find that NMDA nonlinearities of L5 PFC neurons are exploited by the structured connectivity profile in order to implement multiple flexible memoranda with the form of robust network activity trajectories. These trajectories maintain WM information in a dynamic way at the single neuron level, while also exhibiting the representation shift between the stimulus and memory epochs and maintain robust memoranda over time at the population level.

Results

The PFC network model reproduces single neuron and population dynamics.

We developed a compartmental model network of the PFC, able to reproduce both the dynamic single neuron firing and the representational shift from the stimulus to the post-stimulus epoch at the population level. The network consisted of 250 biophysically detailed and morphologically simplified L5 PFC pyramidal neurons (PC) and 83 fast spiking interneurons (FS). Single neuron models were validated against experimental data with respect to both active and passive electrophysiological properties (Table S1). Basal dendrites of pyramidal neurons supported NMDA-dependent non-linear integration, as found experimentally (Figure 1A, Table S2). The PFC network was described by a non-random connectivity matrix, in congruence with available experimental data (Figure S1, Tables S3, 4). Specifically, we set the PC pair connection probability to be (a) distance-dependent and (b) local clustering dependent, based on the common neighbor rule (Perin et al., 2011; Song et al., 2005; Y. Wang et al., 2006). We then rearranged the PC-PC connections to achieve the high reciprocity ratio (47% of total PC-to-PC connections, given that there is a connection) found specifically in the PFC (Morishima et al., 2011; Y. Wang et al., 2006). FS interneurons were connected to PCs in a sparse and distance-dependent way (Packer & Yuste, 2011), respecting the connectivity reciprocity ratio for FS to pyramidal projections in layer 5 frontal cortex (Otsuka & Kawaguchi, 2009) (Figure S1F-H). In addition, 66% of the FS interneurons were connected with each other as reported in (Galarreta & Hestrin, 1999).

We stimulated the model network by delivering 1-sec long, uncorrelated, Poisson-like spike trains to random subpopulations consisting of 50 PCs (20% of total, n=10 trials, Figure 1B and 1C). We focused on the emergence and dynamics of activity following the end of stimulus presentation

(hereby termed post-stimulus epoch), considering all PCs in the network and not just the stimulated ones. Trials with no PC activity up until the last second of the simulation were not considered to induce the memoranda and were not included in the analysis. Both stimulated and non-stimulated neurons were recruited during the post-stimulus epoch and displayed diverse spiking profiles (Figure 1C). The irregularity of PC firing was further increased during the post-stimulus epoch (mean $CV_{stimulus}$=0.42±0.25, mean $CV_{post\text{-}stimulus}$= 0.64±0.46, p value=0.0001, Mann Whitney U test, Figure 1D). Inter-spike intervals (ISIs) were also increased during the post-stimulus epoch ($ISI_{stimulus}$=40.3±53.8ms, $ISI_{post\text{-}stimulus}$= 52.9±115.3, p value=0.0004, Mann Whitney U test).

At the network level, activity was sparse (sparsity=0.9, values close to 1 correspond to sparser activity) and characterized by temporally dynamic population responses; the firing of some neurons persisted, increased or fluctuated during the post-stimulus period (Figure 2A). We measured the correlation coefficient of the binned network activity (bin size q=50ms) between different simulation time points. Figure 2B depicts the de-correlation in network activity during the stimulus and the memory epochs, denoting the shift from the sensory to the memory representation(s). Within the stimulus epoch, population activity is consistently highly correlated, yet, during the memory epoch correlation is stronger over the diagonal time axis, indicating that population activity is not static during the memory epoch. The same shift was observed for the energy state: it peaks after stimulus withdrawal and reaches lower values during the post-stimulus period (Figure 2C). Accordingly, the population velocity drops upon stimulus withdrawal and remains at stable values during the post-stimulus period (Figure 2C). These results replicate 1) the irregular spiking activity of single neurons reported in (Bernacchia et al., 2011; Brody et al., 2003; Jun et al., 2010; Meyers, Freedman, Kreiman, Miller, & Poggio, 2008; Shafi et al., 2007) 2) the de-correlation of the network activity in the late post-stimulus epoch from the stimulus epoch

reported in (Barak et al., 2010; Murray et al., 2017; Stokes et al., 2013) and 3) the low energy state during the post-stimulus epoch reported in (Stokes et al., 2013). Thus, our model network reproduces the complexity of single neuron and population responses as well as the representational shift during the post-stimulus epoch.

A recurrent PFC model network expresses multiple memoranda

The shift from the stimulus-driven network activity to a different, memory-related one is more prominently observed in PCA state space (Figure 3A). Network responses for different trials, when projected to the low-dimensional space, diverge from the stimulus epoch (green) and follow different trajectories in the PCA state space over time. Interestingly, during the late memory epochs (Figure 3A, darker colors, showing 9 trials with activity in post-stimulus epoch), activity appears to converge for some trials to specific trajectories in the PCA state space. In order to quantify the number of distinct memory trajectories over the last second of the post-stimulus period that emerge in the above simulations, we used k-means. The evaluation of the clustering results was done using the Bayesian Information Criterion (BIC, see Methods). We found that for the specific instance of the structured network shown in Figure 3, population activity converged to three different robust trajectories (BIC 213.87). These findings suggest that, additionally to the representational shift from the sensory to the memory epoch, a dynamic mechanism is recruited by specific stimuli in an otherwise fixed connectivity network, which allows its rapid re-configuration, in the absence of any learning mechanism, and amounts to the retention of different memoranda.

NMDA nonlinearities underlie multiple memoranda

We hypothesized that a thresholding mechanism acting in short time scales may engage the network into the distinct robust trajectories. The most prominent such candidate, in the absence of any short- or long- term plasticity, is the NMDA receptor. To evaluate this hypothesis, we repeated the above simulations while eliminating the NMDA conductances (conductance set to zero). To compensate for the reduction in excitability, we increased the AMPA conductance such that the saturated response matched the one when both NMDA and AMPA conductances were included (Figure S2A). We further adjusted the network inhibition (STAR Methods) to ensure the induction of post-stimulus activity (Figure 3B). We found that the AMPA-only model network was unable to create more than one robust trajectory (Figure 3B, BIC=213.2). This was also accompanied by a decrease in the network's sparsity (mean of control sparsity=0.89, mean of AMPA-only sparsity=0.49) and CV (mean of control $CV_{post\text{-}stimulus}$=0.64±0.46, mean of AMPA-only $CV_{post\text{-}stimulus}$= 0.28±0.21, p value=0.0001, Mann Whitney U test) and an increase in network's firing frequency (mean of control $ISI_{post\text{-}timulus}$=52.9±115.3ms, mean of AMPA-only $ISI_{post\text{-}stimulus}$= 16.6±23.6, p value=0.0001, Mann Whitney U test) (Figure S2B).

The NMDA conductance can create two distinct dendritic phenotypes with increasing synaptic magnitude: a) a non-linear jump in the depolarization, due to the Mg^{+2} mediated non-linearity of the NMDA activation function and b) a depolarization plateau potential following the jump, supported by its slow kinetics (Doron, Chindemi, Muller, Markram, & Segev, 2017). We thus performed an additional manipulation to dissect the contribution of the non-linear jump. Specifically, we removed the voltage dependence of the NMDA conductance associated with the Mg^{+2} block, while keeping the slow kinetics and the plateau potential intact (see Figure 2E, Fig S2A,C). Again, the network failed to support more than one robust trajectory (Figure 3C,

BIC=212.9). Note that in this case sparsity is maintained (mean sparsity=0.91) and the ISI/CV change (mean of control $CV_{post\text{-}stimulus}$=0.64±0.46, mean of Mg^{+2} block $CV_{post\text{-}stimulus}$= 0.55±0.33, p value=0.0001, Mann Whitney U test, mean of control $ISI_{post\text{-}timulus}$=52.9±115.3ms, mean of Mg^{+2} block $ISI_{post\text{-}stimulus}$= 50.1±91.8, p value=0.0001, Mann Whitney U test) (Figure S2C). This shows that indeed NMDA receptors mediate the temporal dynamics during the post-stimulus period and act as a thresholding mechanism for the emergence of multiple states.

Synaptic arrangement along the dendrite regulates the network re-configurability

We further assessed the role of the basal dendrites in the NMDA-depended generation of multiple trajectories, since the spatial arrangement along the dendrites influences synaptic integration in non-linear ways (Losonczy & Magee, 2006; Poirazi, Brannon, & Mel, 2003a, 2003b; Polsky, Mel, & Schiller, 2004). We evaluated how many distinct trajectories can be supported when varying the dendritic locations of recurrent synapses between PC pairs and of dendritic inhibitory inputs (n=10). These manipulations resulted in a total of 100 simulated conditions: 10 trials X 10 synaptic arrangements for a given connectivity matrix.

An example of the population activity trajectories for different synaptic arrangements, given a specific connectivity matrix, is shown in Figure 3D. In line with our hypothesis that dendritic integration is a key player, responses in the PCA space vary considerably when changing the synaptic arrangement, giving rise to several distinct activity robust trajectories. Note that these different trajectories emerge in response to stimulation of the same neurons in a given network (fixed connectivity matrix) with the exact same stimuli. The only difference between trials is the dendritic location of the recurrent excitatory connections and inhibitory ones, indicating that dendritic integration is a key player in the observed phenotype.

The different synaptic arrangements could support the same or different trajectories; in other words, given two conditions that support the same number of trajectories, are these identical in the PCA state space? To evaluate this we pooled the responses from the 10 synaptic arrangements and the 10 trials (Figure 4B). This revealed that the different synaptic arrangements induce qualitatively different trajectories. This indicated that for a given connectivity matrix and input, the arrangement of recurrent synapses enables the network to engage in different subsets of PCA state space effectively expanding the flexibility of the network.

We further investigated the generalization of our results by identifying the number of trajectories found for different connectivity matrices (n=4) and per synaptic arrangements (n=10, thus amounting to 40 cases). We found that the probability of a network with a given connectivity and synapse arrangement to express more than a single activity robust trajectory was 0.625 (Figure 4C, blue).

Structured connectivity exploits NMDA nonlinearities to increase memory

Given the above results, we sought to identify whether structural connectivity can selectively exploit the NMDA non-linearities to induce more than one memory robust trajectory. To assess the role of connectivity (both in terms of recurrent connections and the clustering coefficient) in the network's ability to hold multiple memoranda, we ran the same simulations when each PC pair was independently connected with a fixed probability (4 random connectivity matrices, for 10 recurrent synapse arrangements and 10 trials). Both the structured and the random networks exhibited the same average connection probability (~0.12), to ensure a fair comparison. In the presence of the NMDA conductance, the random network produced fewer robust trajectories compared to the structured one (mean of control $CV_{post\text{-}stimulus}$=0.64±0.46, mean of random $CV_{post\text{-}}$

$_{stimulus}$= 0.56±0.42, p value=0.0001, Mann Whitney U test, mean of control $ISI_{post-timulus}$=52.9±115.3ms, mean of random $ISI_{post-stimulus}$= 34.7±77.6, p value=0.0001, Mann Whitney U test) (Figure 4F, Figure S2D). The probability of expressing more than one memory trajectory was 0.375. These findings indicate that structured connectivity can better exploit NMDA nonlinearities for increased network memory dimensionality.

Robust trajectories correspond to different, dynamically recruited neuronal assemblies

We showed that the same network model can robustly encode, multiple memoranda, during the post-stimulus period, utilizing a dynamic thresholding mechanism. How is this mechanism changing the effective connectivity between neurons, to create distinct activity patterns? Are these associated with specific neuronal assemblies and, if yes, how are these assemblies recruited during the post-stimulus period? To answer these questions, we utilized the non-negative matrix factorization (NNMF) approach (D. D. Lee & Seung, 1999), which allows us to approximate any neuronal activity pattern over time as a linear combination of a "basis" set. This "basis" set consists of neuronal activity profiles realized in the network, as opposed to PCA components which do not directly map onto specific neurons (Figure 5A, B). Therefore, resulting factor matrices contain coherent parts of the original network activation enabling both active assemblies and their activity amplitude to be monitored. We applied the NNMF on the network responses for the same network connectivity matrix and synaptic arrangement as used for Figure 3 for different trials. Population activity can be decomposed to the sum of activities of different neuronal assemblies, whereby the contribution of each assembly has a different weight (i.e. assembly activity) (Figure 5C). Note that assembly recruitment can also be transient, yet each trial can be identified by its assembly composition. To identify whether different combinations of assemblies

map to different robust trajectories, we visualized the activity for different trials, color-coded by the trajectory, where same trajectories appear to be represented by the same assembly combinations (Figure 5D). The engagement of the assemblies in a trajectory can change over time, yet their combination appears distinct between the different trajectories. Correlation analysis of number of trajectories and optimal number of assemblies, for the structured network connectivity, shows interesting results. First, that there is a correlation ($r=0.37$ $p=0.017$) of optimal number of assemblies with optimal number of robust trajectories and there are not fewer assemblies than trajectories (Figure 5E). This reveals that trajectories/memories stored in the network tend to be represented by different populations, rather than different firing activation of the same population i.e. stimulus selectivity of the network is mixed both in terms of firing activity and population, rather than same neurons that vary their activity for multiple stimuli in a linear manner (Figure 5D). This is ideal for linearly separating these representations down-stream in complex cognitive tasks. Second, there is not always a one-to-one correspondence of number of robust trajectories and assemblies. This is shown by the identification of more assemblies than robust trajectories for a given network instance (Figure 5E). This observation can be interpreted as having multiple assemblies, encoding a single robust trajectory (Figure 5D, red colored trajectory), and this is interesting, since it could mean that there are assemblies that are reusable across different memoranda and not necessarily linked with representing just a single memorandum. These results suggest that PFC networks, incorporating NMDA-dependent synaptic integration and structured connectivity, result in activity expressed by different active neuronal assemblies, thus increased dimensionality of the memory epoch state space and affecting also the overall compositionality of a single network, during working memory tasks.

DISCUSSION

It is not well understood how information entering the prefrontal cortex (PFC) can be utilized for flexible decision making: how does a single network can represent and support distinct tasks in a rapid, flexible manner (Mante et al., 2013)? Experimental evidence indicates that network connectivity can sustain changes in synaptic efficacies faster than classical plasticity, enabling it to rapidly switch between different activity states, based on the input (Fujisawa et al., 2008; Stokes et al., 2013). To this direction we evaluated the hypothesis that nonlinear dendritic activation and integration, specifically the NMDA conductance, plays an important role in this dynamic network connectivity (Arnsten, Paspalas, Gamo, Yang, & Wang, 2010; Stokes, 2015).

Towards this goal we used a state of the art modeling approach that enables us to dissect the essential mechanisms that drive these highly dynamic phenomena. We built and validated a PFC network model against experimental data that reproduces a large number of experimental observations that are key in understanding dynamic coding in the PFC. More specifically single neuron responses are irregular, yet together create a coherent low dimensional trajectory. Network activity shifts from encoding the noisy stimulus, towards a low energy state that represents the robust memorandum for the specific task at hand, something first time replicated by a biophysical model.

This allowed us to pose key questions regarding the mechanisms that support this behavior. We show that a mechanism responsible for intra-trial, fast, but temporary changes in dynamic connectivity is the NMDA nonlinear integration. This mechanism can rapidly shift synaptic efficacies and enable the network to differentially integrate information for short periods of time (Stokes, 2015). Maximal exploitation of NMDA non-linearities is achieved by connecting subsets

of neurons in a clustered, hyper-reciprocal manner, as evidenced in the PFC anatomy (Y. Wang et al., 2006). The above mechanisms/conditions enable the emergence of functionally connected neuronal assemblies allowing the robust, yet flexible encoding of multiple memoranda in a not per se optimized network model, during WM simulations. During this post-stimulus period patterns of activity of single neurons appear as variable, yet across the network activity assumes a smooth and stable form, with variability mostly vanished and near orthogonal to a linear readout, downstream mechanism, maintaining the separation of the memoranda, as reported experimentally (Murray et al., 2017).

Our findings are consistent with a new experimental view of neurons exhibiting persistent activity of strong temporal dynamics and heterogeneity, that yet retain stability and robustness of WM (Murray et al., 2017)(Meyers et al. 2008; Stokes et al. 2013, Miller, Erickson, and Desimone 1996, J Duncan 2001; John Duncan and Miller 2002). Most importantly, and to our knowledge, this is one of the first studies that show how dendritic integration can be exploited by the structured connectivity profile, already shown to be beneficial (Brunel, 2016). There are several lines of evidence that indicate that NMDA receptors are important for PFC function (M. J. Wang et al., 2013).

Neuronal mixed selectivity does not contrast the classical single neuron selectivity, with neurons able to display gradings of mixed selectivity to stimuli. Yet these experimental observations point away from the belief: single neuron = single variable, thus increasing the capacity and signal to noise ratio (Fusi, Miller, & Rigotti, 2016) of the network, as the distributed specialization encoding over many neurons helps scale memoranda in a single network, yet doesn't induce completely redundant/overlapped responses. This evidence, generalized, adheres to a concept of neuronal ensemble compositionality, where different (specialized) neuronal ensembles of similarly

‘mixed/tuned’ single neurons can be utilized together, as reusable memory units. This compositionality principle can be time saving in acquiring new tasks, since the connectivity does not need to be modified completely, just find/activate a new combination of already learned components, something incredibly important for the computational capabilities of the PFC, given its immediate behavioral importance (X.-J. Wang, Song, Yang, Joglekar, & Newsome, 2019). Our work contributes towards this direction, showing that such a configurability can indeed be supported in a biophysical model network of the PFC and such a clustered configuration is indeed optimal in terms of WM storage, contrasted with an unspecific, random projection between neurons, that compromises the compositionality of the network (Barak, Rigotti, et al., 2013).

Although this work strives to simulate as many variables as possible, the degrees of freedom in biophysical networks are very high, leading eventually to the simulation of the very prominent ones only. We took great care in simulating the extensive connectivity patterns found in experimental PFC works and the diverse pyramidal dendritic apparatus, both excitatory and inhibitory, that we claim to make such a difference. Also, the NMDA is targeted as the mechanism able to shape PFC output. It is definite that this mechanism can be altered by a multitude of parameters, essentially shaping its responses (Arnsten et al., 2010). Although in this work we address the PFC network as a whole, so we abstract NMDA activation over the network response. The specific mechanisms that can alter its performance it is a thought for later work, aiming at further validating this schema. Though the above are just the most prominent characteristics of the network. Multiple number of dendritic inputs per neuron, as well as a more detailed circuit with diverse inhibitory population, might constitute a continuation of this work. Yet more complexity might need different approaches to analyze and reason about the results.

This work also points out the NMDA conductance as a necessary mechanism to dynamic network

connectivity. This can also explain vulnerabilities to cognitive dysfunction with the same mechanism. It is found that PFC connectivity state can be altered with advancing age, genetic insults, during stress or psychiatric disorders (M. Wang et al., 2007). Yet this is something that might not be visible in vitro/in vivo, unless the network is activated in a cognitive operation. Our work enables us to focus on the target mechanism, NMDAr conductance, in order to shed light in the magnitude of is performance in best/worst case scenario with results aligned with experimental predictions (less performance/memory) (M. Wang et al., 2007). This is something difficult to be addressed completely in an experimental way, yet.

Star Methods

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Software and Algorithms		
Algorithms and analysis	This paper	https://github.com/stamatiad/prefrontal_analysis
Model of Pyramidal neuron	Model DB	accession number: 155057
Model of Fast Spiking Interneuron	Model DB	accession number: 168310

CONTACT FOR REAGENT AND RESOURCE SHARING

Further requests for resources should be directed to and will be fulfilled by Stefanos Stamatiadis (stamatiad.st@gmail.com).

METHOD DETAILS

DETAILS OF MODEL NEURONS

Here we provide the details of the model neurons.

The Pyramidal Neuron Model (PC)

The model of the PC was adapted from (Papoutsi, Sidiropoulou, Cutsuridis, & Poirazi, 2013) (Model DB accession number: 155057). In brief, it consisted of five compartments: the soma, the axon, the basal dendrite, the proximal apical dendrite and the distal apical dendrite (apical tuft). The somatic, proximal and distal apical dendritic compartments included Hodgkin–Huxley-type transient and persistent Na^+ currents, voltage-dependent K^+ currents (delayed-rectifier, A-type and D-type) , a fast and slow Ca^{2+} -dependent K^+ current, a hyperpolarization-activated non-specific cation current (h current), and the N-, R-, L- and T- type voltage-dependent Ca^{2+}currents (I_{caN}; I_{caR}; I_{caL}; I_{CaT}). The basal dendrite included transient and persistent Na^+ currents, a delayed K^+ rectifier current, an A-type K^+ current, a D-type K^+ current, a N-type Ca^{2+} current and a h current. The axon included a transient Na^+ current and a delayed rectifier K^+ current. Indicative electrophysiological properties of the PC model are shown in Table S1. For detailed description see (Papoutsi et al., 2013, 2014). We scaled membrane capacitance on dendrites to account for spine additional membrane area.

Fast-Spiking interneuron model (FS)

The FS interneuron model was adapted from (Konstantoudaki, Papoutsi, Chalkiadaki, Poirazi, & Sidiropoulou, 2014) (Model DB accession number: 168310). In brief, the FS interneuron model consisted of a somatic, a dendritic and an axonic compartment. The somatic compartment included mechanisms for the slow A-Type K^+ current, the N-type high-threshold activated Ca^{2+} current and h current. Indicative electrophysiological properties of the FS model are shown in Table S1. For detailed description see (Konstantoudaki et al., 2014).

Synaptic Properties

The AMPA and NMDA currents of PCs were tuned to match the experimental data of (Nevian, Larkum, Polsky, & Schiller, 2007), as per (Papoutsi et al., 2013). The AMPA and NMDA EPSC kinetics of the FS interneuron model were constrained against data from fast spiking layer 5 PFC interneurons of (H.-X. Wang & Gao, 2009), as per (Konstantoudaki et al., 2014). The conductance of the $GABA_A$ mechanism was set so that the amplitude of the mIPSC was 10 pA at a holding potential of -65mV, as per (Woo et al., 2007). Finally, we implemented the slow inhibitory synaptic current ($iGABA_B$). The respective synaptic parameters are listed in Table S2 and detailed description of the synaptic models have been published in (Konstantoudaki et al., 2014; Papoutsi et al., 2013).

DETAILS OF NETWORK CONNECTIVITY

Here we describe the details of connecting model neurons, both in terms of the connectivity matrix as well as the specific somatic/dendritic projection patterns implemented. Vectors are denoted by lowercase bold letters, matrices by uppercase bold letters.

Network Connectivity

The network model consisted of 250 PCs (75%) and 83 FSs (25%) (Beaulieu, 1993; Dombrowski, Hilgetag, & Barbas, 2001). We implemented an experimentally constrained connectivity by treating the biophysical network as a directed graph and its connectivity as a square, binary, connectivity matrix $\mathbf{A} \in \mathbb{R}^{(M+F)}$**,** where M is the number of PCs and F is the number of FSs. In this matrix, the rows represented neurons' outgoing connections (efferents) and the columns their incoming connections (afferents), for each neuronal type respectively, with

PCs preceding FSs (for connectivity matrix index $i, i \leq M$ corresponds to PCs and $M < i \leq M + F$ corresponds to FSs). Although this matrix contains the connectivity for the whole network, for brevity we will refer to the first M by M part of **A**, as its *excitatory* part, since it contains the PC-PC connectivity only. We implemented two different connectivity configurations between the PCs: one highly non-random (structured) and one random. In the structured configuration, a) PC pair connection probability was distance dependent, b) almost half connections were reciprocals, as reported in the PFC (Y. Wang et al., 2006) and c) the PC pair connection probability was re-adjusted to be local-clustering dependent, based on experimental data (Perin, Telefont, & Markram, 2013; Song et al., 2005). In the random configuration, each pair was connected independently with a fixed probability. Both types of configurations exhibited the same average probability ($P_{average}$) of connection (~0.12), measured as described below.

In the structured configuration, to implement the distance dependent connectivity constraints, somata of PCs were regarded as points randomly scattered in a 180x180x180 μm^3 cubic space (Figure S1A) using a method similar to (Perin et al., 2011), resulting in a somatic density of $6x10^4$neurons/mm^3 (Schuz & Palm, 1989). Distances between these points were calculated by wrapping the cube dimensions in order to eliminate edge effects (Packer & Yuste, 2011; Perin et al., 2011). Each unique unordered PC pair {i, j} (from now on referred to as pair), could be connected in two ways: either a) unidirectionally (i → j or $j \rightarrow i$) or b) reciprocally (i ↔ j). These disjoint events followed the experimentally reported probabilities as a function of distance in layer 5 (Perin et al., 2011). In order to reproduce the hyper-reciprocity reported in the PFC, we set the ratio of the reciprocal and unidirectional connections to be approximately 1. Specifically, we implemented the following probability density functions (pdf):

$$g^{overall}(x) = 0.22\ \mathrm{e}^{-0.0052\ \mathrm{x}}\ (1)$$

$$g^{reciprocal}(x) = 0.12\ \mathrm{e}^{-0.0085\,\mathrm{x}}\ (2)$$

with $g^{overall}(\mathrm{x})$ being the pdf of having any connection and $g^{reciprocal}(x)$ is the pdf of having reciprocal connection, both as a function of the distance x. We derived the unidirectional function from their difference,

$$g^{unidirectional}(\mathrm{x}) = g^{overall}(\mathrm{x}) - \ g^{reciprocal}(x)\ (3)$$

with $g^{unidirectional}(x)$ being the pdf of having non reciprocal connection as a function of the distance x (Figure S1B). Following, for each pair its reciprocal or unidirectional probability of connection, $\mathrm{P}_{\{i,j\}}^{reciprocal}$ and $\mathrm{P}_{\{i,j\}}^{unidirectional}$ respectively was:

$$\mathrm{P}_{\{i,j\}}^{unidirectional} \equiv\ P((i \to j) \cup (j \to i)) - P((i \to j) \cap (j \to i)) = g^{unidirectional}(\ \mathrm{d(i,j)}\ )\ (4)$$

$$\mathrm{P}_{\{i,j\}}^{reciprocal} \equiv\ P((i \to j) \cap (j \to i)) = g^{reciprocal}\big(\ \mathrm{d(i,j)}\big)(5)$$

where, $\mathrm{d(i,j)}$ is the pair's intersomatic distance. For each pair the probabilities sum up to the unit:

$$1 = \mathrm{P}_{\{i,j\}}^{unidirectional} + \mathrm{P}_{\{i,j\}}^{reciprocal} + \left(1 - \mathrm{P}_{\{i,j\}}^{overall}\right)(6)$$

The excitatory part of the connectivity matrix **A** was created by rejection sampling for each pair. Specifically, we sampled from a uniform random distribution once for each pair and the excitatory part of the connectivity matrix **A** was determined as:

$$\begin{gathered}
a_{i,j} = 1, a_{j,i} = 1, if\ \mathrm{r}_{\{i,j\}} < \mathrm{P}_{\{i,j\}}^{reciprocal} \\
a_{i,j} = 1, a_{j,i} = 0, if\ \mathrm{P}_{\{i,j\}}^{reciprocal} \leq \mathrm{r}_{\{i,j\}} < \mathrm{P}_{\{i,j\}}^{reciprocal} + \frac{\mathrm{P}_{\{i,j\}}^{unidirectional}}{2} \\
a_{i,j} = 0, a_{j,i} = 1, if\ \mathrm{P}_{\{i,j\}}^{reciprocal} + \frac{\mathrm{P}_{\{i,j\}}^{unidirectional}}{2} \leq \mathrm{r}_{\{i,j\}} < \mathrm{P}_{\{i,j\}}^{reciprocal} + \mathrm{P}_{\{i,j\}}^{unidirectional} \\
a_{i,j} = 0, a_{j,i} = 0, if\ \mathrm{P}_{\{i,j\}}^{reciprocal} + \mathrm{P}_{\{i,j\}}^{unidirectional} \leq \mathrm{r}_{\{i,j\}} < 1
\end{gathered} \tag{7}$$

where $\mathrm{r}_{\{i,j\}}$ is a random number on the interval [0,1] and $a_{i,j}$ the elements of $\boldsymbol{A}$ matrix. The resulting $\mathrm{P}_{\mathrm{average}}$ was ~0.12.

To implement the local-clustering dependent constrains we used the common neighbor rule using an iterative approach. According to this rule, pairs sharing increasing number of common neighbors are more probable to be connected (Perin et al., 2011, 2013). A common neighbor is defined as a third PC that receives from or projects to a PC pair. Starting from the above distance-dependent excitatory part of connectivity matrix $\boldsymbol{A}$ we iteratively recreated the connectivity to reproduce the common neighbor rule. Specifically, in each iteration we calculated the number of common neighbors for all pairs of the network, creating the symmetric matrix $\mathbf{C_N}$.

$$\mathbf{C_N} = \mathrm{E} \times \mathrm{E}^T \quad (8)$$

, with E was the undirected (and thus symmetric) binary, connectivity matrix of the graph of the excitatory part of $\mathbf{A}$. Following, we normalized the matrix of common neighbors

$$\widehat{\mathbf{C_N}} = \frac{\mathbf{C_N}}{\|\mathbf{C_N}\|_{max}} \quad (9)$$

so that the pair of neurons sharing the maximal number of common neighbors is assigned a factor of connection $\widehat{\mathbf{C_N}}_{i,j} = 1$ and pairs sharing no common neighbors a factor of $\widehat{\mathbf{C_N}}_{i,j} = 0$ (Perin et al., 2013). The updated connection probability of a pair ($\mathrm{P}^{cn}_{\{i,j\}}$) is its $\mathrm{P}^{overall}_{\{i,j\}}$, multiplied with $\widehat{\mathbf{C}_{\boldsymbol{N}}}$.

$$\mathrm{P}^{cn}_{\{i,j\}} = \mathrm{P}^{overall}_{\{i,j\}}\, \widehat{\mathbf{C}\text{-}\mathbf{N}}_{i,j} (10)$$

Then, an updated overall connection probability is calculated from the $\mathrm{P}^{cn}_{\{i,j\}}$ so that the $\mathrm{P}_{\mathrm{average}}$ remains constant:

$$\mathrm{P}^{updated\ overall}_{\{i,j\}} = \mathrm{P}^{cn}_{\{i,j\}} \frac{\sum_{i\in\{1,\ldots,M\},j\in\{1,\ldots,M\},i\neq j} \mathrm{P}^{overall}_{\{i,j\}}}{\sum_{i\in\{1,\ldots,M\},j\in\{1,\ldots,M\},i\neq j} \mathrm{P}^{cn}_{\{i,j\}}} \quad (11)$$

This updated overall connection probability is then used to calculate the unidirectional and reciprocal connection probabilities. Towards this, we calculated the distance estimate $\mathrm{d}'(\mathrm{i},\mathrm{j})$,

based on the $\mathrm{P}_{\{i,j\}}^{updated\ overall}$ for each pair. The unidirectional (and respective the reciprocal) connection probabilities were then calculated using this distance estimate and the excitatory part of $\boldsymbol{A}$, was recreated with rejection sampling as above.

$$\mathrm{P}_{\{i,j\}}^{unidirectional} \equiv P((i \to j) \cup (j \to i)) - P((i \to j) \cap (j \to i)) = g^{unidirectional}(\ \mathrm{d}'(\mathrm{i},\mathrm{j})\ ) \quad (12)$$

$$\mathrm{P}_{\{i,j\}}^{reciprocal} \equiv P((i \to j) \cap (j \to i)) = g^{reciprocal}\Big(\ \mathrm{d}'(\mathrm{i},\mathrm{j})\Big) \quad (13)$$

This algorithm ended after 1000 iterations, where the average clustering coefficient of the network plateaus as in (Perin et al., 2011) (Figure S1C). The clustering coefficient (Watts & Strogatz, 1998) for a single PC was defined as

$$\mathrm{C_m} = \frac{k_m^{available}}{k_m(k_m - 1)/2}, m = 1, \dots, M \quad (14)$$

where k_m is the number of neighbors (undirected edges to other neurons) that neuron m have and $k_m^{available}$ are the (undirected) edges that exist between these neighbors.

In the case of the randomly connected network, pyramidal neurons were connected with the same independent probability for each pair, irrespective of distance, computed so that the two networks (structured and random) have the same $\mathrm{P_{average}}$:

$$\mathrm{P}_{\{i,j\}}^{random} = (\sum_{\mathrm{i}=1}^{\mathrm{M}} \sum_{\mathrm{j}=1}^{\mathrm{M}} \mathbf{A}_{\boldsymbol{i,j}}\ )\ /\ (\mathrm{M}^2 - \mathrm{M}) \quad (15)$$

where $\mathbf{A}$ is the excitatory part of the connectivity matrix of the structured configuration, M is the number of pyramidal cells of the network. The respective connectivity matrix for the random configuration was created by rejection sampling.

We stimulated a total of n=4 differently network instances for each case (4 structured and 4 random network instances), characterized by their respective connectivity matrix. These

connectivity instances were then translated in the synaptic connectivity of the detailed biophysical models respecting, existing anatomical and neurophysiological data. Connectivity between PFC FSs and PCs was made in a dense, non-selective manner and in a distance dependent way (Packer & Yuste, 2011) as reported experimentally. Due to the lack of data for the distance-dependent connectivity of L5 PFC FS interneurons with nearby PCs we used the distance dependent pdf between L5 somatosensory neurons from (Packer & Yuste, 2011). Connectivity properties are summarized in Table S4. The connectivity between FSs and PCs is contained in connectivity matrix **A**. Specifically, we began by defining a FS-PC overall, distance dependent connection pdf:

$$g^{FS-PCoverall}(x) = \mathrm{e}^{-\mathrm{x}/180} \quad (16)$$

similar to what is reported for somatosensory L5 (Packer & Yuste, 2011). Then we proceeded in creating reciprocal and unidirectional pdfs, based on the FS-PC overall pdf:

$$g^{FS-PCreciprocal}(x) = g^{FS-PCoverall}(x)\ 0.283 \quad (17)$$

$$g^{PC-FS}(x) = g^{FS-PCoverall}(x)\ 0.035 \quad (18)$$

$$g^{FS-PC}(x) = g^{FS-PCoverall}(x)\ 0.014 \quad (19)$$

respecting the percentages reported in (Otsuka & Kawaguchi, 2009). Because the overall connection probability reported was conflicting (Otsuka & Kawaguchi, 2009; Packer & Yuste, 2011), we averaged the overall FS-PC connection probability to 0.46, with the above pdfs summing up to this probability for zero distance. The PC-FS connections were added to the connectivity matrix **A** by rejection sampling for each pair, as previously, where the pair $\{i, j\}$ now refers to FS-PC neurons ($M * F$ pairs in total). Specifically, we sampled from a uniform random distribution once for each pair and the respective elements of the connectivity matrix, $a_{i,j}$ were determined as:

$$a_{i,j} = 1, a_{j,i} = 1, if\ \mathrm{r}_{\{i,j\}} < \mathrm{P}_{\{i,j\}}^{FS-PCreciprocal}$$
$$a_{i,j} = 1, a_{j,i} = 0, if\ \mathrm{P}_{\{i,j\}}^{FS-PCreciprocal} \leq \mathrm{r}_{\{i,j\}} < \mathrm{P}_{\{i,j\}}^{FS-PCreciprocal} + \mathrm{P}_{\{i,j\}}^{FS-PC}$$
$$a_{i,j} = 0, a_{j,i} = 1, if\ \mathrm{P}_{\{i,j\}}^{FS-PCreciprocal} + \mathrm{P}_{\{i,j\}}^{FS-PC} \leq \mathrm{r}_{\{i,j\}} < \mathrm{P}_{\{i,j\}}^{FS-PCreciprocal} + \mathrm{P}_{\{i,j\}}^{FS-PC} + \mathrm{P}_{\{i,j\}}^{PC-FS}$$
$$a_{i,j} = 0, a_{j,i} = 0, if\ \mathrm{P}_{\{i,j\}}^{FS-PCreciprocal} + \mathrm{P}_{\{i,j\}}^{FS-PC} + \mathrm{P}_{\{i,j\}}^{PC-FS} \leq \mathrm{r}_{\{i,j\}} < 1 \tag{20}$$

where $\mathrm{r}_{\{i,j\}}$ is a random number on the interval [0,1] and $a_{i,j}$ the elements of $\boldsymbol{A}$ matrix. FS-PC connectivity elements of **A**, lie on its part with index $i, M < i \leq M + F$ and $j, j \leq M$ with i corresponding to FS index and j corresponding to PC index and its symmetric (indices can be used interchangeably between neuronal types, since the pairs are unordered).

Finally, the connectivity between FSs is contained in connectivity matrix **A**, along with previous connections. For this neuronal type no distance dependent pdfs were used. Specifically, we set the connection probabilities as: $\mathrm{P}_{\{i,j\}}^{FS-FSreciprocal} = 0.045$ and $\mathrm{P}_{\{i,j\}}^{FS-FSunidirectional} = 0.068$, resulting in 0.18 overall connection probability as reported in (Galarreta & Hestrin, 1999) (Table S4, Figure S1). FS to FS connections were added to the connectivity matrix **A** by rejection sampling for each pair, as previously, where the pair $\{i, j\}$ now refers to FS-FS neurons, with $(F^2 - F)/2$ pairs in total. Specifically, we sampled from a uniform random distribution once for each pair and the respective elements of the connectivity matrix, $a_{i,j}$ were determined as:

$$a_{i,j} = 1, a_{j,i} = 1, if\ \mathrm{r}_{\{i,j\}} < \mathrm{P}_{\{i,j\}}^{FS-FSreciprocal}$$
$$a_{i,j} = 1, a_{j,i} = 0, if\ \mathrm{P}_{\{i,j\}}^{FS-FSreciprocal} \leq \mathrm{r}_{\{i,j\}} < \mathrm{P}_{\{i,j\}}^{FS-FSreciprocal} + \frac{\mathrm{P}_{\{i,j\}}^{FS-FSunidirectional}}{2}$$
$$a_{i,j} = 0, a_{j,i} = 1, if\ \mathrm{P}_{\{i,j\}}^{FS-FSreciprocal} + \frac{\mathrm{P}_{\{i,j\}}^{FS-FSunidirectional}}{2} \leq \mathrm{r}_{\{i,j\}} < \mathrm{P}_{\{i,j\}}^{FS-FSreciprocal} + \mathrm{P}_{\{i,j\}}^{FS-FSunidirectional}$$
$$a_{i,j} = 0, a_{j,i} = 0, if\ \mathrm{P}_{\{i,j\}}^{FS-FSreciprocal} + \mathrm{P}_{\{i,j\}}^{FS-FSunidirectional} \leq \mathrm{r}_{\{i,j\}} < 1 \tag{21}$$

where $\mathrm{r}_{\{i,j\}}$ is a random number on the interval [0,1] and $a_{i,j}$ the elements of $\boldsymbol{A}$ matrix. FS-FS connectivity elements of **A**, lie on its last F-by-F part with index $i, M < i \leq M + F$ and $j, M <$

$j \leq M + F$ with indices corresponding to FS interchangeably, since the pairs are unordered.

Synaptic connectivity properties

The number and percentage of synapses per type (inhibitory/excitatory) of connection is summarized in Table S3 and Table S4 respectively. Individual neuron connectivity properties including the location and the number of synaptic contacts, as well as the latencies between pairs of neurons, were based on anatomical and electrophysiological data. Specifically, the recurrent connections between pyramidal neurons were at the basal dendrite through 5 synapses, as reported in (Markram, Lübke, Frotscher, Roth, & Sakmann, 1997). Synaptic latencies were drawn from a Gaussian distribution with $\mu = 1.7$ms and $\sigma = 0.9$ according to (Thomson & Lamy, 2007). The precise location of the connecting synapses was random with uniform probability across the basal dendrite of each pyramidal neuron. These locations were constant across trials, but reshuffled when the network model underwent a synaptic rearrangement. PCs connected to FSs via 2 synaptic contacts (Buhl et al., 1997), with the corresponding latencies drawn from a Gaussian distribution with $\mu = 0.6$ms and $\sigma = 0.2$ (Angulo, Rossier, & Audinat, 1999). FSs project to PCs both to the soma and the basal dendrites as reported in (Kubota et al., 2015). Their location was varied randomly with uniform probability between the synaptic rearrangements. The conductance was calculated by exponentially interpolating the reported values in (Kubota et al., 2015) (Table S2, Figure S3). FS interneurons project to FS with a single GABAa synapse of 1.7nS (TableS2/S3) as in (Galarreta & Hestrin, 1999). Synaptic latencies were drawn from a Gaussian distribution with $\mu = 1.76$ms and $\sigma = 0.07$ according to (Bacci, Huguenard, & Prince, 2003).

NETWORK STIMULATION AND MANIPULATION

Here we describe the stimulation protocols and any manipulations made to the default synaptic mechanism state described above.

Stimulation

In each trial we stimulated 50 randomly selected (with uniform probability) PCs with uncorrelated Poisson inputs (λ=0.06, resulting in about 60Hz afferent spiking input) for 5sec. The incoming stimulus (40 AMPA and NMDA excitatory synapses), was provided to the proximal apical dendrites of the selected PCs (Kuroda, Yokofujita, & Murakami, 1998). For each connectivity matrix (n=4), 10 synaptic arrangements (defined as randomly positioning the excitatory recurrent PC-PC connections and inhibitory connections along the basal dendrite) were implemented, thus resulting in a total of 40 cases. We repeated for a total of T=10 times the stimulation (trials, h) of each of the network case, whereby the 50 stimulated PCs and the stimulus timings were varied. We did not include in the analysis the trials with no activity from any pyramidal neuron in the last second of the post-stimulus epoch.

Synaptic manipulations

In order for the network to produce post-stimulus epoch activity we made compensations to the overall recurrent excitation and feedback inhibition, by scaling respectfully the synaptic conductances. Specifically, for the structured configuration a baseline was achieved by augmenting the recurrent excitation (g_{NMDA} and g_{AMPA}) with a x1.75 factor and the feedback inhibition (g_{GABAa} and g_{GABAb}) with a factor of x3. The following compensations are derived from this baseline. To compensate for changes in excitability in simulations where g_{NMDA} conductance was set to zero, the g_{AMPA} was augmented 50 times to saturate to the same somatic EPSP with increasing the number of pair-pulse inputs at 50Hz (Figure S2A). Feedback inhibition (g_{GABAa} and g_{GABAb}) was adjusted to one third, in order to allow for post-stimulus activity.

We further manipulated the properties of the NMDA receptors. Specifically for the g_{NMDA} , it was simulated as in (Polsky, Mel, & Schiller, 2009)

$$\mathrm{I_{NMDA}} = g(V - E_{rev})(22)$$

$$\mathrm{I_{Ca^{+2}}} = \frac{\mathrm{I_{NMDA}}}{10}(23)$$

$$\mathrm{g}_{NMDA} = \frac{\mathrm{A} - \mathrm{B}}{1 + ne^{-gamma\,v}}(24)$$

$$\frac{\mathrm{dA}}{dt} = \frac{-\mathrm{A}}{tau_a}(25)$$

$$\frac{\mathrm{dB}}{dt} = \frac{-\mathrm{B}}{tau_b}(26)$$

, where $E_{rev} = 0mV$, $tau_a = 90\ ms$, $tau_b = 5\ ms$, $gamma = 0.08\ mV^{-1}$, $n = 0.25mM^{-1}$.

In order to remove the characteristic non-linear 'jump' from the PCs, we removed the exponent from equation (24) that simulated Mg^{2+} blockage:

$$\mathrm{g_{NMDA}} = (\mathrm{A} - \mathrm{B})(27)$$

resulting in equation (27) and the saturating depolarization response of Figure S2B. No other compensations of excitation or inhibition were needed in this case to observe post-stimulus period activity.

In simulations with the random network connectivity we adjusted the inhibition to half of the baseline, for the network to produce post-stimulus period activity.

NETWORK ANALYSIS

Here we describe how we proceed in annotating and analyzing network activity. We start by defining network activity states and metrics. We continue with grouping these activity states in robust trajectories and with how we can detect active assemblies in the network, along with each

method validation practices. Vectors are denoted by lowercase bold letters, matrices by uppercase bold letters.

Network states

Population activity was considered as discrete vectors composed of neuronal activity (number of spikes) binned across time in a temporal window q, tracing trajectories in the neuronal activity state space. We constructed the instantaneous network state (INS) $\boldsymbol{u} \in \mathbb{R}^M$ using the binned activity of each $m = 1, \dots, M$ pyramidal neurons. We used non-overlapping windows q = 50ms and this resulted in $M = simulation\ time\ (ms)/q$ INSs. For a given trial h, we calculated the neuronal activity state space trajectories of the N INSs, resulting in a matrix $\boldsymbol{B}^h = [\boldsymbol{u}_1, \boldsymbol{u}_2, \dots, \boldsymbol{u}_N], \boldsymbol{B}^h \in \mathbb{R}^{M \times N}$. We analyzed the population activity, across the trials $h = 1, \dots, T$, thus resulting in a total of $L = T * N$ INSs. This resulted in a M -by- L matrix of the concatenated INS, $\boldsymbol{D}^i = \left[\boldsymbol{B}^1, \boldsymbol{B}^1, \dots, \boldsymbol{B}^T\right], \boldsymbol{D}^i \in \mathbb{R}^{M \times L}$. The latter is referred to as **network instance state matrix** and characterized the activity of each network instance (same network connectivity and synaptic arrangement, different trials).

Cross-correlation

Cross correlation between INSs was calculated using the Pearson correlation

$$\boldsymbol{R}_{ij} = \frac{\boldsymbol{S}_{ij}}{\sqrt{\boldsymbol{S}_{ii}\boldsymbol{S}_{jj}}}, \forall i, j \in N \tag{28}$$

where $\boldsymbol{S}_{ij}$ is the covariance of INS $\vec{\boldsymbol{u}}_i$ with INS $\vec{\boldsymbol{u}}_j$.

Sparsity

To assess the sparseness of network response we used the measure proposed by (Rolls, 1997) subtracted from unity, so that higher values correspond to more sparse activity:

$$S_{TR} = 1 - \frac{\left(\sum_m \frac{\hat{r}_m}{M}\right)^2}{\sum_m \frac{\hat{r}_m{}^2}{M}}, \forall m \in M \tag{29}$$

Where $\hat{r}_m$ is the average firing rate of neuron $m \in \mathbb{R}^M$.

Population multidimensional velocity

Network multidimensional velocity V_{MD} was calculated as the gradient of the activity with respect to time:

$$V_{MD}(n) = \frac{d(\vec{u}_n, \vec{u}_{n+1})}{\Delta \mathrm{t}}, \mathrm{n} = 1, \dots, \mathrm{N} \tag{30}$$

Where $d(\vec{u}_n, \vec{u}_{n+1})$ is the Euclidean distance in $\mathbb{R}^M$ of the two adjacent in time INS, converted in Hz, for a $\Delta t = 50\ ms$. The per trial average was used.

Population energy velocity

We define the vector containing the average firing frequency of the neurons as the energy vector of the network, per INS. We define the rate of change of this vector as the *Energy Velocity* V_E i.e. the change in the overall energy level of the system, across the duration of one trial.

$$V_E(n) = \sum_1^M r_m^n, n = 1, \dots, N \tag{31}$$

Where r_m is the firing rate of neuron $m \in \mathbb{R}^M$ in state n.

Principal component analysis

INSs was transformed to *instantaneous PCA network states* (IPS) using PCA to examine the neuronal dynamics in the lower dimensional space. The *network instance state matrix* $\boldsymbol{D}^i$ was transformed with PCA to $\boldsymbol{T}^i$, $\boldsymbol{T}^i \in \mathbb{R}^{M \times L}$. Following we split the L principal component loadings to the respective T trials, representing the *network state space trajectories*, per trial. For plotting,

raw PCA data were used. The prominent principal components were the ones with eigenvalues above unity.

We evaluated the emergence of multiple state space trajectories using the k-means clustering method and validation with the Bayesian Information Criterion (BIC) (Wit, van den Heuvel, & Romeijn, 2012). This is a way to discern the optimal number of discrete robust trajectories, since it amounts in separable network activity by a probabilistic metric for the last second of memory epoch that has $G = \frac{1sec}{q}$ INSs. From the network state space trajectories (matrix $\boldsymbol{T}^i$) we kept the last second of each trial in the matrix $\boldsymbol{L}^h$, $\boldsymbol{L}^h \in \mathbb{R}^{M_{PCA} \times G}, h = 1, \dots, T$, keeping only M_{PCA} principal components as described above. Each trajectory was smoothed prior to clustering with a Savitzky–Golay filter with of window size 31 and of polynomial order 3. We initialized the centroids of k-means using the mean-shift algorithm over all trajectory data (columns of $\boldsymbol{L}^h$) and across all trials. We used mean-shift to locate the 'density points' in network state space i.e. points with more overall activity across all trials. We initialized mean-shift with T Gaussian kernels, that could converge to an equal or smaller number of local maxima. Subsequently we utilized only K of the latter as the k-means initial centroids (with $1 \leq K \leq T$).

Specifically, we compute the mean of each network state space trajectory across time, for each trial

$$\boldsymbol{v}_i^h = \frac{1}{G} \sum_{j=1}^{G} \boldsymbol{L}_{i,j}^h, \forall i, i = 1, \dots, M_{PCA} \tag{32}$$

getting T points on network state space $\boldsymbol{v^h} \in \mathbb{R}^{M_{PCA}}$.

The mean-shift evolve these points iteratively, computing a local average centered to each point with a gaussian kernel. The weighted mean of the density at these points is determined by Gaussian kernel window K_{RBF} as:

$$\bar{\mu}(\boldsymbol{v^h}) = \frac{\sum_{\boldsymbol{q}_j \in Z} K_{RBF}(\boldsymbol{q}_j, \boldsymbol{v^h})\boldsymbol{q}_j}{\sum_{\boldsymbol{q}_j \in Z} K_{RBF}(\boldsymbol{q}_j, \boldsymbol{v^h})} \tag{33}$$

across neighboring IPS $\boldsymbol{q}_j$ (pooled all trajectory points) where Z is the neighborhood of $\boldsymbol{v^h}$**:** a set of points for which $K_{RBF}(\boldsymbol{q}_j, \boldsymbol{v^h}) \leq 0.1$. The K_{RBF} is defined as:

$$K_{RBF}\left(\boldsymbol{q}_j, \boldsymbol{v^h}\right) = e^{\left(-\frac{\left\|\boldsymbol{q}_j - \boldsymbol{v^h}\right\|^2}{2\sigma^2}\right)}, \tag{34}$$

with $\boldsymbol{q}_j$ being a trajectory point close to $\boldsymbol{v^h}$ and setting the variance σ^2 equal to the average variance across runs.

Then the resulting density centroids $\bar{\mu}(\boldsymbol{v^h})$ are used as the new Gaussian kernel centers $\boldsymbol{v^h} := \bar{\mu}(\boldsymbol{v^h})$ in the next iteration. The iteration continues until centers converge to non-moving positions.

Next, mean-shift density centroids taken in pairs were merged, starting from those with minimal Euclidean distance, until only K remained.

The k-means algorithm is initialized from the latter and minimized the average intra-cluster Mahalanobis distance, Because the variance in each PCA was different:

$$D_M(\boldsymbol{x}_h) = \sqrt{(\boldsymbol{x}_h - \boldsymbol{\mu}_k)^T \boldsymbol{S}_k^{-1} (\boldsymbol{x}_h - \boldsymbol{\mu}_k)} \tag{35}$$

where $\boldsymbol{\mu}_k$ is cluster $k = 1, \ldots, K$ centroid, $\boldsymbol{x}_h$ the average activity across time of trial h that belongs to cluster k and $\boldsymbol{S}_k^{-1}$ is the inverse of the covariance matrix of the data of cluster k. K-means used the Euclidean distance between $\boldsymbol{\mu}_k$ and $\boldsymbol{x}_h$ only once, during initialization, to label clusters, since no information about covariance $\boldsymbol{S}_k^{-1}$ for each cluster exist in the first iteration.

Clustering evaluation was done utilizing the Bayesian Information Criterion (BIC):

$$BIC(k) = \ln(T)k - 2\ln(\hat{L}) \tag{36}$$

where $\widehat{L}$ is the value of the log-likelihood function across all T data points (one for each trial), clustered in K clusters.

$$L^h = \frac{e^{\frac{-D_M(\boldsymbol{x_h})}{2}}}{\sqrt{(2\pi)^{M_{PCA}} \det(\boldsymbol{S})}}$$

$$\widehat{L} = \sum_{h=1}^{T} \ln(L^h) \tag{37}$$

where $D_M(\boldsymbol{x_h})$ is the Mahalanobis distance of point $\boldsymbol{x_h}$ from its cluster centroid and $\boldsymbol{S}$ the cluster centroid covariance matrix.

Since minimizing the BIC corresponds to maximizing the posterior model probability for the given observations, we choose the optimal BIC using the model with the smaller BIC value.

$$K^* = \arg\min_{k=1\ldots K} BIC(k) \tag{38}$$

In cases where network activity resulted in a single cluster of network state space trajectories, the k-means algorithm will start partitioning within that single cluster and the above methodology will overfit the BIC model, designating optimal K^*strongly biased from individual run's distribution (and variance). To avoid this overfitting we set an upper bound on K^*, K^*_{max} by setting a lower bound on the inter-cluster distance:

$$K^*_{max} = \arg\min_{k=1\ldots K} \left(d < \min_{k,k' \in \{1\ldots K\}} (\|\boldsymbol{\mu}_k - \boldsymbol{\mu}_{k'}\|_d) \right) \tag{39}$$

where we set the Mahalanobis distance threshold to $d = 3$ between cluster centroids $\boldsymbol{\mu}_k$ and $\boldsymbol{\mu}_{k'}$ and we search the minimum k, avoiding partitioning data further.

Non-Negative Matrix Factorization

To identify the active neuronal population, we used nonnegative matrix factorization (NNMF) as in (D. D. Lee & Seung, 1999). This method utilizes an approximate factorization of the form:

$$\boldsymbol{D}^i \approx \boldsymbol{W^i H^i} \tag{40}$$

were $\boldsymbol{W^i} \in \mathbb{R}^{M\times K}$ and $\boldsymbol{H^i} \in \mathbb{R}^{K\times L}$ matrices across cases $i = 1, \ldots, 40$. The matrix $\boldsymbol{W^i}$ corresponds to K basis patterns (columns of $\boldsymbol{W^i}$, neuronal assemblies, Figure 5A) and the rows of the matrix $\boldsymbol{H^i}$contain the activity of each of the assemblies over the L INSs. Following we split the L assembly activity coefficients to the respective T trials.

The optimal number of assemblies $\mathrm{K}^*_{\boldsymbol{assemblies}}$ was calculated with a K-fold Cross-Validation, across multiple K. The network activity (observation sample) is randomly partitioned into $\mathrm{K}_{CV} = 20$ equally sized subsamples to be used in equal number of K-fold iterations. Of these a single subsample is retained as the *testing data set* for the NNMF (model) and the remaining $\mathrm{K}_{CV} - 1$ subsamples are used as the *training data set*, for each K-fold iteration. Specifically, we divided the network instance state matrix, $\mathbf{D^i}$ in K_{CV} random partitions to be used throughout a single K-fold Cross-Validation. Towards this, we defined a binary mask matrix $\mathbf{O}$, with each of its elements indicating a left out data point $\mathbf{O}^k_{i,j} = 0$ (testing data), or a data point in the training data set $\mathbf{O}^k_{i,j} = 1$, for K-fold iteration k=1,…, K_{CV}. This matrix changes for each K-fold Cross-Validation iteration, as predefined by the initial random partition, reflecting the different testing and training datasets. Following, we calculated the $\mathbf{W}, \mathbf{H}$ for each k. Because with this manipulation the matrix $\boldsymbol{D}^i$ has missing elements, we performed gradient descent to estimate the NNMF factor matrices $\mathbf{W}, \boldsymbol{H}$ as in (D. Lee & Seung, 2001):

Algorithm 1: NNMF with missing elements

1: Initialize W, H randomly

2: **for** t = 1 to 1000 **do**

3: $W_{ij} \leftarrow W_{ij} \frac{((O \circ D)H^T)_{ij}}{((O \circ WH)H^T)_{ij}}$

4: $H_{ij} \leftarrow H_{ij} \frac{(W^T(O \circ D))_{ij}}{(W^T(O \circ WH))_{ij}}$

5: **end for**

where $\circ$ denotes the Hadamard product. We then evaluated the performance (error) of the model using the testing data set in each K-fold iteration k

$$\mathrm{D}_{NNMF}^{k} = \frac{\left\| (\mathbf{1} - \mathbf{O}^{\boldsymbol{k}}) \circ (\boldsymbol{W}^{\boldsymbol{i}} \boldsymbol{H}^{\boldsymbol{i}} - \boldsymbol{D}^{i}) \right\|_{\mathrm{F}}}{J} \quad (41)$$

where $(\mathbf{1} - \mathbf{O}^{\boldsymbol{k}})$ selects the testing data, $\boldsymbol{W}^{\boldsymbol{i}} \boldsymbol{H}^{\boldsymbol{iT}}$ are the NNMF factor matrices of the training data and M and L are the number of rows and columns of $\boldsymbol{D}^{i}$ respectively, for K-fold iteration k. The error is normalized with J that is the number of zero elements of $\mathbf{O}^{\boldsymbol{k}}$ when calculated for the testing data and the number of its nonzero elements when calculated for the training data. After all K-fold iterations are completed, we take the mean error as the Cross-Validation error for the NNMF with complexity K. Calculating this error for different number of assemblies, K creates a 'bias-variance' type of error curve, with its minimum indicating the optimal number of assemblies that **D** can be divided, avoiding overfitting i.e. choosing large K for the NNMF model. Due to noise from the partition procedure, the cross-validation procedure was repeated n=50 times for different K (number of assemblies). We define the optimal number of assemblies as the value of K where D_{NNMF}^{k} is minimum.

$$K_{assemblies}^{*} = \arg \min_{k=1 \ldots K} \left(\mathrm{D}_{NNMF}^{k} \right) \quad (42)$$

DATA AND SOFTWARE AVAILABILITY

The microcircuit model was implemented in the NEURON (version 7.4) simulation environment

(Hines & Carnevale, 2001) and simulations were executed on a 312 core cluster, exploiting the parallelization offered for the multiple cores. Software was written in the Python (http://python.org) programming language. Implementations of algorithms used to compute quantities presented in this study are available at: https://github.com/stamatiad/prefrontal_analysis.

Acknowledgements

Author Contributions

Conceptualization, S.S.S., P.A., P.P.; Data curation, S.S.S., P.A.; Formal analysis, S.S.S.; Funding acquisition, P.P.; Investigation, S.S.S.; Methodology, S.S.S.; Project administration, P.P.; Resources, P.P.; Software, S.S.S; Supervision, P.A., P.P.; Validation, S.S.S.; Visualization, P.A., P.P.; Writing – original draft, S.S.S, P.A., P.P.

Declaration of Interests

The authors declare no competing interests.

References

Angulo, M. C., Rossier, J., & Audinat, E. (1999). Postsynaptic glutamate receptors and integrative properties of fast-spiking interneurons in the rat neocortex. *Journal of Neurophysiology*, *82*(3), 1295–1302.

Arnsten, A. F. T., Paspalas, C. D., Gamo, N. J., Yang, Y., & Wang, M. (2010). Dynamic network connectivity: A new form of neuroplasticity. *Trends in Cognitive Sciences*, *14*(8), 365–375. https://doi.org/10.1016/j.tics.2010.05.003

Bacci, A., Huguenard, J., & Prince, D. (2003). Functional autaptic neurotransmission in fast-spiking interneurons: a novel form of feedback inhibition in the neocortex. *The Journal of*

Neuroscience, *23*(3), 859–866. Retrieved from http://www.jneurosci.org/content/23/3/859.short

Barak, O., Rigotti, M., & Fusi, S. (2013). The Sparseness of Mixed Selectivity Neurons Controls the Generalization-Discrimination Trade-Off. *Journal of Neuroscience*, *33*(9), 3844–3856. https://doi.org/10.1523/JNEUROSCI.2753-12.2013

Barak, O., Sussillo, D., Romo, R., Tsodyks, M., & Abbott, L. F. (2013). From fixed points to chaos: Three models of delayed discrimination. *Progress in Neurobiology*, *103*, 214–222. https://doi.org/10.1016/j.pneurobio.2013.02.002

Barak, O., Tsodyks, M., & Romo, R. (2010). Neuronal population coding of parametric working memory. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *30*(28), 9424–9430. https://doi.org/10.1523/JNEUROSCI.1875-10.2010

Beaulieu, C. (1993). Numerical data on neocortical neurons in adult rat, with special reference to the GABA population. *Brain Research*, *609*(1–2), 284–292. https://doi.org/10.1016/0006-8993(93)90884-P

Bernacchia, A., Seo, H., Lee, D., & Wang, X.-J. (2011). A reservoir of time constants for memory traces in cortical neurons. *Nature Neuroscience*, *14*(3), 366–372. https://doi.org/10.1038/nn.2752

Bouchacourt, F., & Buschman, T. (2019). A flexible model of working memory, 1–41. https://doi.org/10.32470/ccn.2018.1231-0

Brody, C. D., Hernández, A., Zainos, A., & Romo, R. (2003). Timing and Neural Encoding of Somatosensory Parametric Working Memory in Macaque Prefrontal Cortex. *Cerebral Cortex*, *13*(11), 1196–1207. https://doi.org/10.1093/cercor/bhg100

Brunel, N. (2016). Is cortical connectivity optimized for storing information? *Nature*

Neuroscience, *19*(5). https://doi.org/10.1038/nn.4286

Buhl, E. H., Tamás, G., Szilágyi, T., Stricker, C., Paulsen, O., & Somogyi, P. (1997). Effect, number and location of synapses made by single pyramidal cells onto aspiny interneurones of cat visual cortex. *The Journal of Physiology*, *500 (Pt 3*, 689–713.

Buonomano, D. V, & Maass, W. (2009). State-dependent computations: spatiotemporal processing in cortical networks. *Nature Reviews. Neuroscience*, *10*(2), 113–125. https://doi.org/10.1038/nrn2558

Dombrowski, S. M., Hilgetag, C. C., & Barbas, H. (2001). Quantitative architecture distinguishes prefrontal cortical systems in the rhesus monkey. *Cerebral Cortex (New York, N.Y. : 1991)*, *11*(10), 975–988.

Doron, M., Chindemi, G., Muller, E., Markram, H., & Segev, I. (2017). Timed Synaptic Inhibition Shapes NMDA Spikes, Influencing Local Dendritic Processing and Global I/O Properties of Cortical Neurons. *Cell Reports*, *21*(6), 1550–1561. https://doi.org/10.1016/j.celrep.2017.10.035

Durstewitz, D., Seamans, J. K., & Sejnowski, T. J. (2000). Neurocomputational models of working memory. *Nature Neuroscience*, *3 Suppl*(november), 1184–1191. https://doi.org/10.1038/81460

Fujisawa, S., Amarasingham, A., Harrison, M. T., & Buzsáki, G. (2008). Behavior-dependent short-term assembly dynamics in the medial prefrontal cortex. *Nature Neuroscience*, *11*(7), 823–833. https://doi.org/10.1038/nn.2134

Funahashi, S., Bruce, C. J., & Goldman-Rakic, P. S. (1989). Mnemonic coding of visual space in the monkey's dorsolateral prefrontal cortex. *Journal of Neurophysiology*, *61*(2), 331–349.

Fusi, S., Miller, E. K., & Rigotti, M. (2016). Why neurons mix: High dimensionality for higher

cognition. *Current Opinion in Neurobiology*, *37*, 66–74. https://doi.org/10.1016/j.conb.2016.01.010

Fuster, J. M., & Alexander, G. E. (1971). Neuron activity related to short-term memory. *Science*, *173*(997), 652–654.

Galarreta, M., & Hestrin, S. (1999). A network of fast-spiking cells in the neocortex connected by electrical synapses. *Nature*, *402*(6757), 72–75. https://doi.org/10.1038/47029

Hines, M. L., & Carnevale, N. T. (2001). NEURON: a tool for neuroscientists. *The Neuroscientist : A Review Journal Bringing Neurobiology, Neurology and Psychiatry*, *7*(2), 123–135. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11496923

Jun, J. K., Miller, P., Hernández, A., Zainos, A., Lemus, L., Brody, C. D., & Romo, R. (2010). Heterogenous population coding of a short-term memory and decision task. *The Journal of Neuroscience*, *30*(3), 916–929. https://doi.org/10.1523/JNEUROSCI.2062-09.2010

Konstantoudaki, X., Papoutsi, A., Chalkiadaki, K., Poirazi, P., & Sidiropoulou, K. (2014). Modulatory effects of inhibition on persistent activity in a cortical microcircuit model. *Frontiers in Neural Circuits*, *8*(January), 7. https://doi.org/10.3389/fncir.2014.00007

Kubota, Y., Kondo, S., Nomura, M., Hatada, S., Yamaguchi, N., Mohamed, A. A., … Kawaguchi, Y. (2015). Functional effects of distinct innervation styles of pyramidal cells by fast spiking cortical interneurons. *ELife*, (July 2015), 1–27. https://doi.org/10.7554/eLife.07919

Kuroda, M., Yokofujita, J., & Murakami, K. (1998). An ultrastructural study of the neural circuit between the prefrontal cortex and the mediodorsal nucleus of the thalamus. *Progress in Neurobiology*, *54*(4), 417–458. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9522395

Lee, D. D., & Seung, H. S. (1999). Learning the parts of objects by non-negative matrix factorization. *Nature*, *401*(6755), 788–791. https://doi.org/10.1038/44565

Lee, D., & Seung, S. (2001). Algorithms for Non-negative Matrix Factorization. *Advances in Neural Information Processing Systems 13*, (1), 556–562.

Losonczy, A., & Magee, J. C. (2006). Integrative properties of radial oblique dendrites in hippocampal CA1 pyramidal neurons. *Neuron*, *50*(2), 291–307. https://doi.org/10.1016/j.neuron.2006.03.016

Mante, V., Sussillo, D., Shenoy, K. V, & Newsome, W. T. (2013). Context-dependent computation by recurrent dynamics in prefrontal cortex. *Nature*, *503*(V). https://doi.org/10.1038/nature12742

Markram, H., Lübke, J., Frotscher, M., Roth, A., & Sakmann, B. (1997). Physiology and anatomy of synaptic connections between thick tufted pyramidal neurones in the developing rat neocortex. *The Journal of Physiology*, *500 (Pt 2*, 409–440.

Meyers, E. M., Freedman, D. J., Kreiman, G., Miller, E. K., & Poggio, T. (2008). Dynamic population coding of category information in inferior temporal and prefrontal cortex. *Journal of Neurophysiology*, *100*(3), 1407–1419. https://doi.org/10.1152/jn.90248.2008

Morishima, M., & Kawaguchi, Y. (2006). Recurrent connection patterns of corticostriatal pyramidal cells in frontal cortex. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *26*(16), 4394–4405. https://doi.org/10.1523/JNEUROSCI.0252-06.2006

Morishima, M., Morita, K., Kubota, Y., & Kawaguchi, Y. (2011). Highly differentiated projection-specific cortical subnetworks. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *31*(28), 10380–10391.

https://doi.org/10.1523/JNEUROSCI.0772-11.2011

Murray, J. D., Bernacchia, A., Roy, N. A., Constantinidis, C., Romo, R., & Wang, X.-J. (2017). Stable population coding for working memory coexists with heterogeneous neural dynamics in prefrontal cortex. *Proceedings of the National Academy of Sciences*, *114*(2), 394–399. https://doi.org/10.1073/pnas.1619449114

Nevian, T., Larkum, M. E., Polsky, A., & Schiller, J. (2007). Properties of basal dendrites of layer 5 pyramidal neurons: a direct patch-clamp recording study. *Nature Neuroscience*, *10*(2), 206–214. https://doi.org/10.1038/nn1826

Otsuka, T., & Kawaguchi, Y. (2009). Cortical inhibitory cell types differentially form intralaminar and interlaminar subnetworks with excitatory neurons. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *29*(34), 10533–10540. https://doi.org/10.1523/JNEUROSCI.2219-09.2009

Packer, A. M., & Yuste, R. (2011). Dense, unspecific connectivity of neocortical parvalbumin-positive interneurons: a canonical microcircuit for inhibition? *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *31*(37), 13260–13271. https://doi.org/10.1523/JNEUROSCI.3131-11.2011

Papoutsi, A., Kastellakis, G., & Poirazi, P. (2017). Basal tree complexity shapes functional pathways in the prefrontal cortex. *Journal of Neurophysiology*, *118*(4), 1970–1983. https://doi.org/10.1152/jn.00099.2017

Papoutsi, A., Sidiropoulou, K., Cutsuridis, V., & Poirazi, P. (2013). Induction and modulation of persistent activity in a layer V PFC microcircuit model. *Frontiers in Neural Circuits*, *7*, 161. https://doi.org/10.3389/fncir.2013.00161

Papoutsi, A., Sidiropoulou, K., & Poirazi, P. (2014). Dendritic Nonlinearities Reduce Network

Size Requirements and Mediate ON and OFF States of Persistent Activity in a PFC Microcircuit Model. *PLoS Computational Biology*, *10*(7), e1003764. https://doi.org/10.1371/journal.pcbi.1003764

Perin, R., Berger, T. K., & Markram, H. (2011). A synaptic organizing principle for cortical neuronal groups. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(13), 5419–5424. https://doi.org/10.1073/pnas.1016051108

Perin, R., Telefont, M., & Markram, H. (2013). Computing the size and number of neuronal clusters in local circuits. *Frontiers in Neuroanatomy*, *7*(February), 1. https://doi.org/10.3389/fnana.2013.00001

Poirazi, P., Brannon, T., & Mel, B. W. (2003a). Arithmetic of subthreshold synaptic summation in a model CA1 pyramidal cell. *Neuron*, *37*(6), 977–987. https://doi.org/10.1016/S0896-6273(03)00148-X

Poirazi, P., Brannon, T., & Mel, B. W. (2003b). Pyramidal neuron as two-layer neural network. *Neuron*, *37*(6), 989–999. https://doi.org/10.1016/S0896-6273(03)00149-1

Polsky, A., Mel, B. W., & Schiller, J. (2004). Computational subunits in thin dendrites of pyramidal cells. *Nature Neuroscience*, *7*(6), 621–627. https://doi.org/10.1038/nn1253

Polsky, A., Mel, B. W., & Schiller, J. (2009). Encoding and decoding bursts by NMDA spikes in basal dendrites of layer 5 pyramidal neurons. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *29*(38), 11891–11903. https://doi.org/10.1523/JNEUROSCI.5250-08.2009

Renart, A., Song, P., & Wang, X.-J. (2003). Robust spatial working memory through homeostatic synaptic scaling in heterogeneous cortical networks. *Neuron*, *38*, 473–485. Retrieved from http://www.sciencedirect.com/science/article/pii/S0896627303002551

Rigotti, M., Barak, O., Warden, M. R., Wang, X.-J., Daw, N. D., Miller, E. K., & Fusi, S. (2013). The importance of mixed selectivity in complex cognitive tasks. *Nature*, *497*(7451), 585–590. https://doi.org/10.1038/nature12160

Rolls, T. T. (1997). Neural Networks and Brain Function, 428. Retrieved from http://book.douban.com/subject/11752244/

Schuz, A., & Palm, G. (1989). Density of neurons and synapses in the cerebral cortex of the mouse. *Journal of Comparative Neurology*, *455*, 442–455.

Shafi, M., Zhou, Y., Quintana, J., Chow, C., Fuster, J., & Bodner, M. (2007). Variability in neuronal activity in primate cortex during working memory tasks. *Neuroscience*, *146*(3), 1082–1108. https://doi.org/10.1016/j.neuroscience.2006.12.072

Song, S., Sjöström, P. J., Reigl, M., Nelson, S. B., & Chklovskii, D. B. (2005). Highly nonrandom features of synaptic connectivity in local cortical circuits. *PLoS Biology*, *3*(3), e68. https://doi.org/10.1371/journal.pbio.0030068

Stokes, M. G. (2015). 'Activity-silent' working memory in prefrontal cortex: a dynamic coding framework. *Trends in Cognitive Sciences*, 1–12. https://doi.org/10.1016/j.tics.2015.05.004

Stokes, M. G., Kusunoki, M., Sigala, N., Nili, H., Gaffan, D., & Duncan, J. (2013). Dynamic coding for cognitive control in prefrontal cortex. *Neuron*, *78*(2), 364–375. https://doi.org/10.1016/j.neuron.2013.01.039

Thomson, A. M., & Lamy, C. (2007). Functional maps of neocortical local circuitry. *Frontiers in Neuroscience*, *1*(1), 19–42. https://doi.org/10.3389/neuro.01.1.1.002.2007

Van Kerkoerle, T., Self, M. W., & Roelfsema, P. R. (2017). Layer-specificity in the effects of attention and working memory on activity in primary visual cortex. *Nature Communications*, *8*, 1–12. https://doi.org/10.1038/ncomms13804

Wang, H.-X., & Gao, W.-J. (2009). Cell type-specific development of NMDA receptors in the interneurons of rat prefrontal cortex. *Neuropsychopharmacology : Official Publication of the American College of Neuropsychopharmacology*, *34*(8), 2028–2040. https://doi.org/10.1038/npp.2009.20

Wang, M. J., Yang, Y., Wang, C.-J., Gamo, N. J., Jin, L. E., Mazer, J. a, … Arnsten, A. F. T. (2013). NMDA receptors subserve persistent neuronal firing during working memory in dorsolateral prefrontal cortex. *Neuron*, *77*(4), 736–749. https://doi.org/10.1016/j.neuron.2012.12.032

Wang, M., Ramos, B. P., Paspalas, C. D., Shu, Y., Simen, A., Duque, A., … Arnsten, A. F. T. (2007). Alpha2A-adrenoceptors strengthen working memory networks by inhibiting cAMP-HCN channel signaling in prefrontal cortex. *Cell*, *129*(2), 397–410. https://doi.org/10.1016/j.cell.2007.03.015

Wang, X.-J. (1999). Synaptic basis of cortical persistent activity: the importance of NMDA receptors to working memory. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *19*(21), 9587–9603.

Wang, X.-J., Song, H. F., Yang, G. R., Joglekar, M. R., & Newsome, W. T. (2019). Task representations in neural networks trained to perform many cognitive tasks. *Nature Neuroscience*, *22*(2), 297–306. https://doi.org/10.1038/s41593-018-0310-2

Wang, Y., Markram, H., Goodman, P. H., Berger, T. K., Ma, J., & Goldman-Rakic, P. S. (2006). Heterogeneity in the pyramidal network of the medial prefrontal cortex. *Nature Neuroscience*, *9*(4), 534–542. https://doi.org/10.1038/nn1670

Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of “small-world” networks. *Nature*, *393*(6684), 440–442. https://doi.org/10.1038/30918

Wimmer, K., Nykamp, D. Q., Constantinidis, C., & Compte, A. (2014). Bump attractor dynamics in prefrontal cortex explains behavioral precision in spatial working memory. *Nature Neuroscience*, (November 2013). https://doi.org/10.1038/nn.3645

Wit, E., van den Heuvel, E., & Romeijn, J. W. (2012). 'All models are wrong. ': An introduction to model uncertainty. *Statistica Neerlandica*, *66*(3), 217–236. https://doi.org/10.1111/j.1467-9574.2012.00530.x

Woo, R.-S., Li, X.-M., Tao, Y., Carpenter-Hyland, E., Huang, Y. Z., Weber, J., … Mei, L. (2007). Neuregulin-1 enhances depolarization-induced GABA release. *Neuron*, *54*(4), 599–610. https://doi.org/10.1016/j.neuron.2007.04.009

Figures

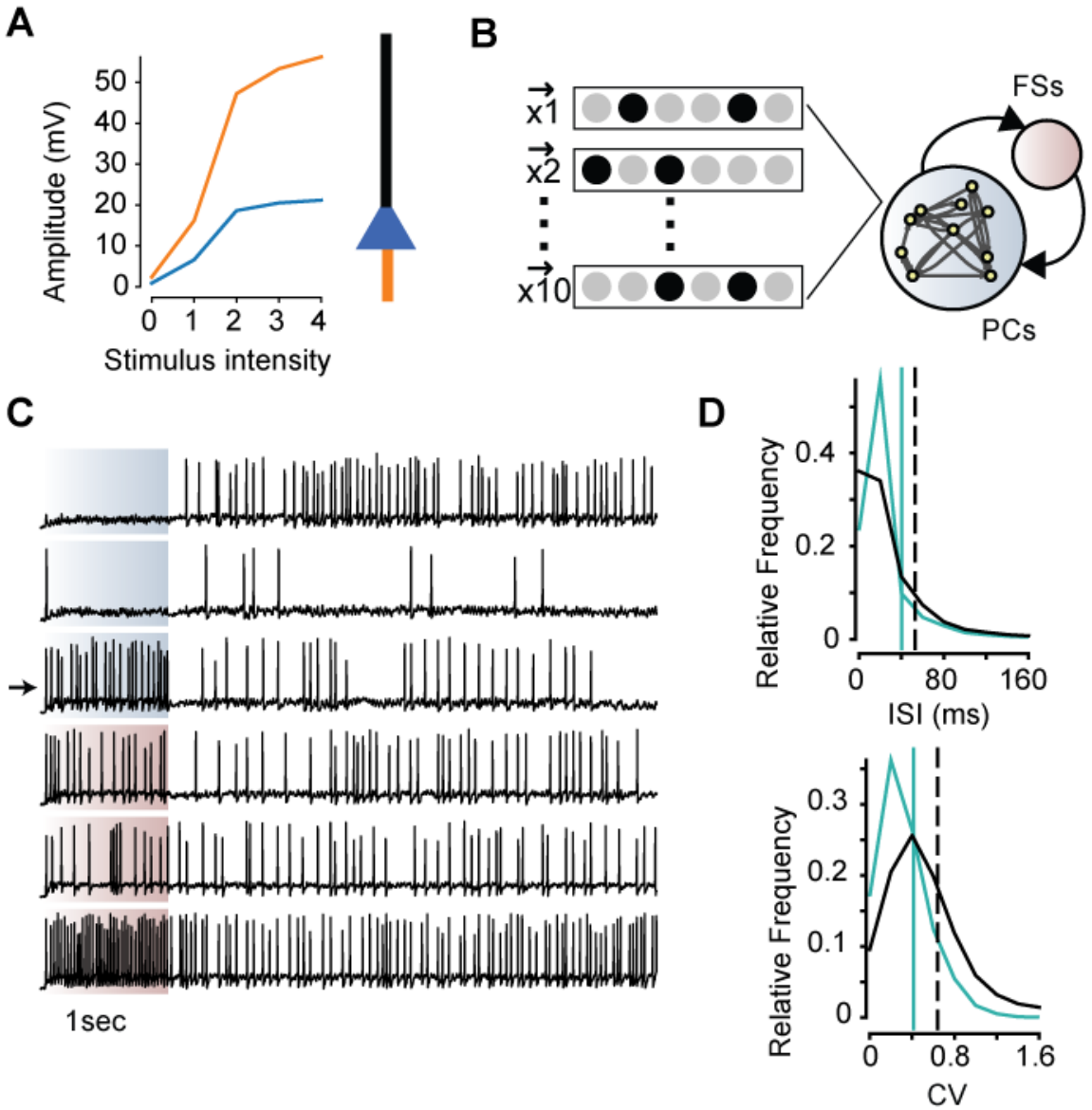


Figure 1 | PFC network model reproduces single neuron dynamics.

(A) Left: Non-linear NMDA responses are generated in the basal dendrites of the pyramidal neurons as in (Nevian et al., 2007). Orange: voltage response of the basal dendrite while increasing the number of activated synapses (2 events at 50Hz, activation of 1 to 21 synapses, with step 5). Blue: Corresponding somatic response. Right: Cartoon morphology of the

pyramidal model. Orange: basal dendritic compartment, blue: somatic compartment, black: apical dendritic compartments (B) Schematic representation of the network model. Fifty randomly selected PCs (n=10 trials) were stimulated. PCs and FSs connectivity profile was based on experimental data (see also Figure S1). (C) Exemplar voltage responses of PCs (top three) and FSs (bottom three) in a single trial. Shaded region indicates stimulus presentation (1sec). Arrow: PC that received stimulation. (D) Histograms of inter-spike intervals (ISIs) (top) and Coefficient of Variation of ISIs (CV, bottom) for PCs. Data are pooled for different structured network instances and trials for the stimulus period (green) and post-stimulus period (black). Distribution means are indicated with vertical lines.

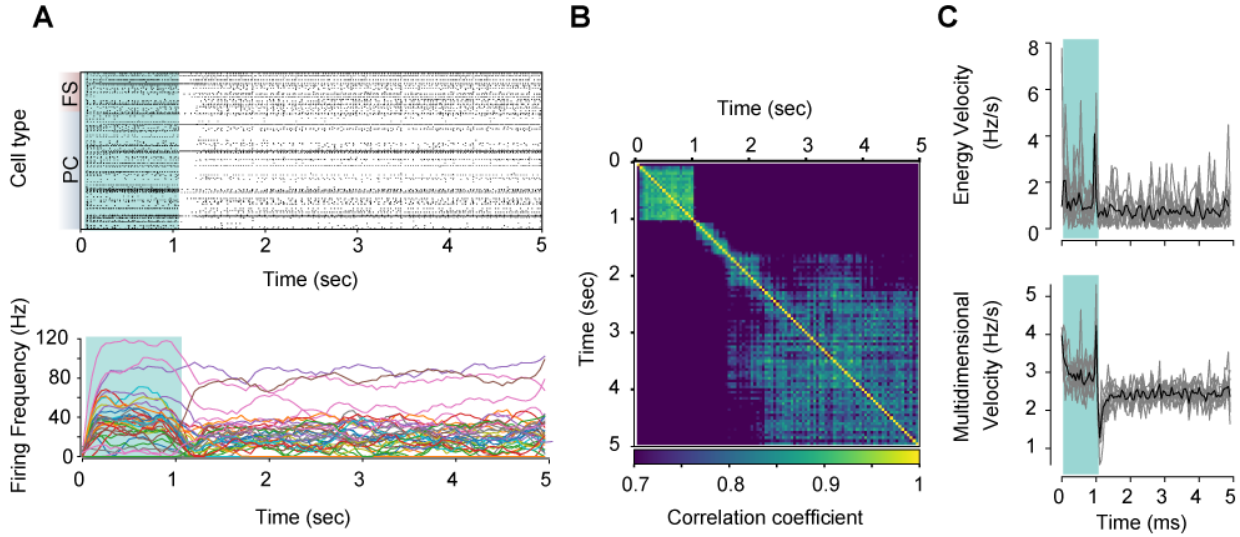
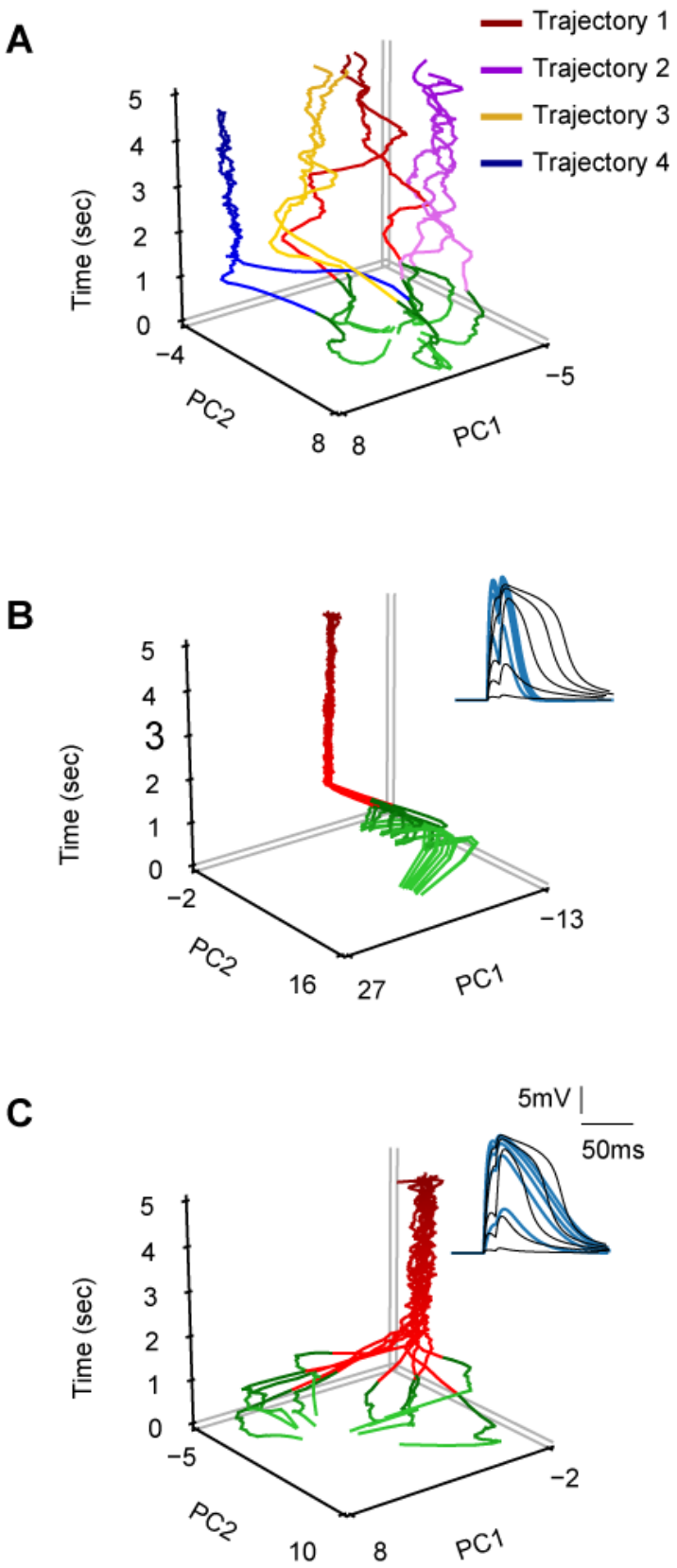


Figure 2 | Dynamic population activity following stimulation.

(A) Top: Raster plot of network activity to a 1 sec stimulus (shaded green area) for pyramidal neurons and interneurons. Bottom: Same trial's smoothed (temporal window: 11ms) firing frequencies of each pyramidal, showing the different and highly dynamic responses over time. Plotted are PCs with average firing frequency across stimulus and post-stimulus epochs above 10Hz. (B) Cross correlation of network states (activity of each pyramidal neuron in a temporal

window q=50ms) across the stimulus and post-stimulus epochs (aligned on stimulus onset, stimulus duration: 1sec). Color bar indicates the correlation coefficient, re-aligned to 0.7-1 range. (C) Network response energy (mean firing rate) velocity, aligned on stimulus epoch onset. (D) Multidimensional velocity of network response, aligned on stimulus epoch onset. C,D: shaded region corresponds to stimulus epoch.


A
Trajectory 1
Trajectory 2
Trajectory 3
Trajectory 4
Time (sec)
5
4
3
2
1
0
−4
PC2
8
8
PC1
−5
B
Time (sec)
5
4
3
2
1
0
−2
PC2
16
27
PC1
−13
C
5mV
50ms
Time (sec)
5
4
3
2
1
0
−5
PC2
10
8
PC1
−2

994

Figure 3 | NMDA nonlinearities create distinct stable trajectories.

(A) Population activity (n=10 trials) reduced in the first two principal components over time. Stimulus epoch is colored green. Different colors (yellow, red, magenta, blue) correspond to different clusters of post-stimulus trajectories, as dictated by k-means analysis. Color map luminosity (brighter to darker) denotes time progression. (B) As before, when removing the NMDA synaptic conductance and compensating by increasing AMPA conductance. This results in the formation of one trajectory for different stimuli. Inset: non-linear response at the soma (black) for the control condition (NMDA and AMPA). Both the non-linear jump and the depolarizing plateau are eliminated (blue), for increasing synaptic drive in the single neuron when NMDA conductance is removed (basal dendrite stimulation, 2 events at 50Hz, activation of 1 to 21 synapses, see also Figure S2A,B). (C) As before, after removal of the Mg^{2+} blockage. This again results in the formation of a single trajectory. Inset: By removing the Mg^{2+} blockage, the non-linear jump is removed (black control, blue Mg^{2+} blockage), while the depolarizing plateau is retained (basal dendrite stimulation, 2 events at 50Hz, activation of 1 to 21 synapses, see also Figure S2A,C).

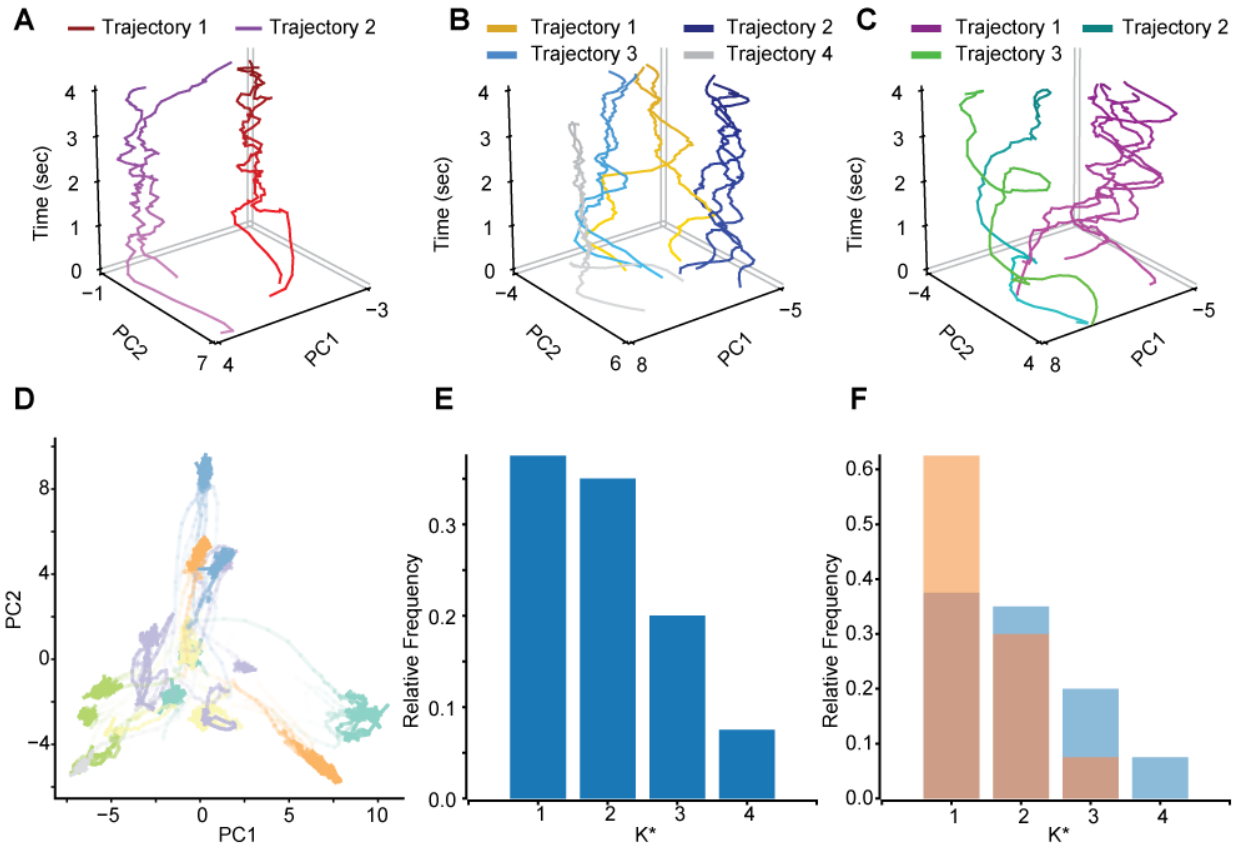


Figure 4 | Network activity states in response to same inputs, showcasing network re-configurability. (A,B,C) Population activity in PCA space for the same structured network instance, but over three synaptic rearrangements. Clusters of trajectories are color coded. It is evident that the number of trajectories produced, changes due to the synaptic rearrangements. The same neurons are stimulated in each trial between rearrangements. (D) Same network instance, but for n=10 synaptic rearrangements, showcasing that trajectories in large part do not overlap (10 trials, 10 synaptic rearrangements). Color code by synaptic rearrangement . (E) Histogram of optimal number of clusters (K*) after k-means for all n=4 structured network instances of n=10 synaptic rearrangements, for n=10 trials each. (F) In the presence of NMDA nonlinear synaptic integration, a randomly connected network produces fewer trajectories compared to the structured one (see also Figure S2D).

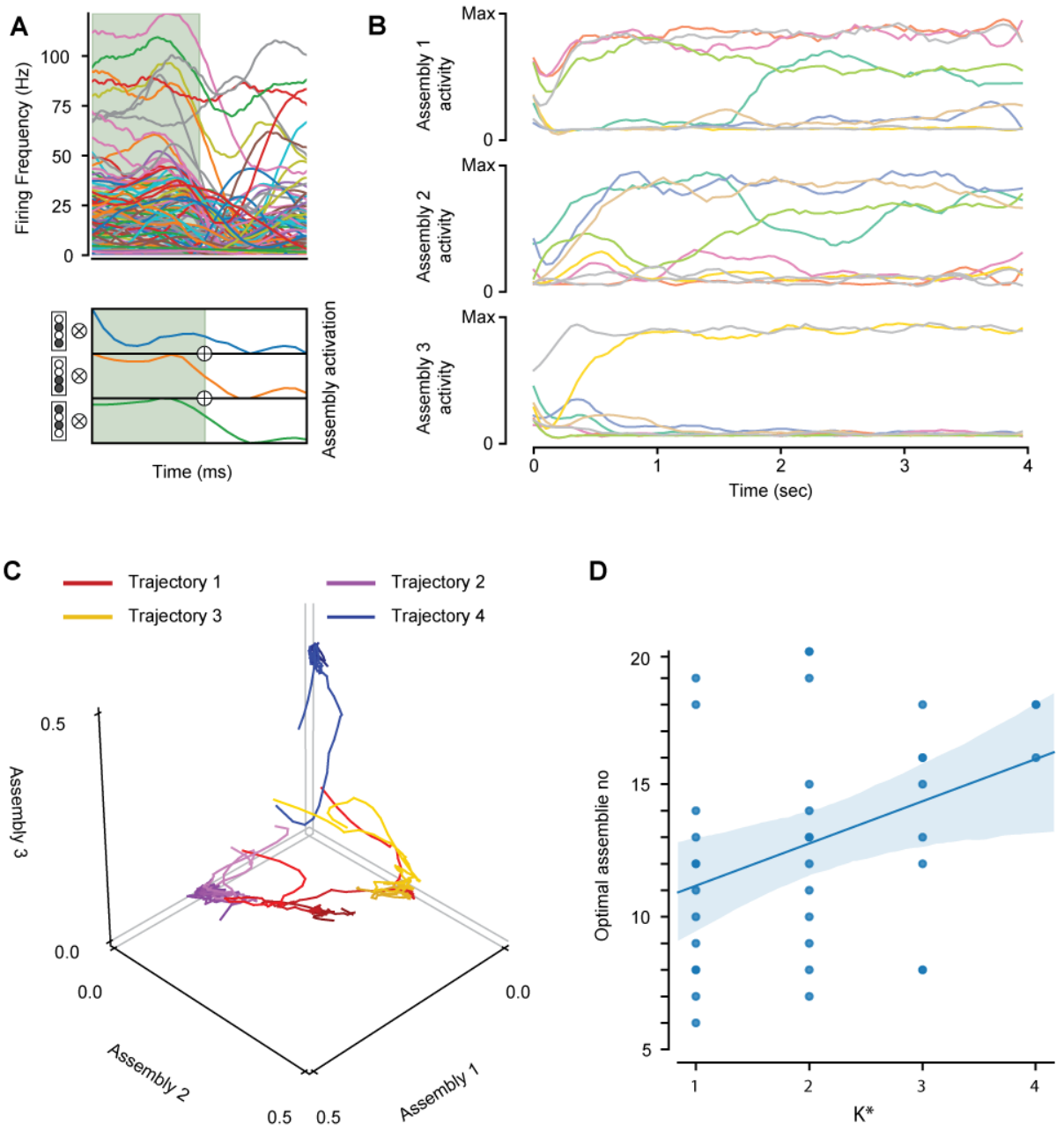


Figure 5 | Different neuronal assemblies are active for each population activity trajectories. (A) Top: Smoothed single neuron activities for the last 0.5 sec of stimulus (green area) and the initial 0.5 sec of the post-stimulus epoch of a given trial. Bottom: Same activity, but now reduced

with NNMF to 3 components (assemblies). The dynamic and highly irregular single neuron activity, appears smooth in the assembly space and the shift in the representation from the stimulus to post-stimulus epoch is evident. The order of operations to go from the assembly space to the original is overlaid. Left: toy examples of assemblies created by 4 neurons. Each neuron can participate in more than one assembly. Each assembly is expressed as a vector of participating neurons. Right: each of the assemblies is multiplied (outer product) with its activity vector over time. Population activity is reconstructed by the summation of the respective matrices across assemblies. (B) Post-stimulus epoch activity in 9 trials (9 out of 10 with PC activity up until the last second of the simulation) reduced with NNMF to three assemblies and presented each time as the activity of each assembly. Note that assembly activity is dynamic across time. Colors correspond to the different trials. (C) Same data as in B, color coded per the trajectories identified in Figure 3A and plotted in the assembly space. It is evident that trajectories that cluster in state space (similar colors), appear to correspond to the activity of specific assemblies/assembly combinations. Note that assemblies are quasi-orthogonal (a neuron can be engaged in more than one assembly) and orthonormal axis are just for visualization purposes. This is expected due to the existence of neurons with mixed selectivity to memoranda. (D) Correlation analysis of optimal number of trajectories and assemblies shows that there is correlation ($r=0.37$ $p=0.017$), yet a single trajectory might not necessarily be created from only one assembly.