**Identification and Targeting of Cortical Ensembles with Probabilistic Graphical Models**

Luis Carrillo-Reid1, Shuting Han1, Tony Jebara2 and Rafael Yuste1.

1 Neuro Technology Center, Department of Biological Sciences, Columbia University, New York, NY 10027, USA

2 Department of Computer Science, Columbia University, New York, NY, 10027, USA

**Corresponding authors:**

Luis Carrillo-Reid. Ph.D. lc2998@columbia.edu

Rafael Yuste, M.D., Ph.D. rafaelyuste@columbia.edu

906 NWC Building 550 West 120 Street, Box 4822, New York, NY, 10027.

**Author Contributions:** L.C.-R & S.H. contributed equally to this work. Conceptualization, L.C.-R, T.J. and R.Y.; Methodology, L.C.-R., S.H. & T.J.; Investigation, L.C.-R.; Writing – Original Draft, L.C.-R. & S.H.; Writing – Review & Editing, L.C.-R., S.H., T.J., & R.Y.; Resources, L.C.-R., S.H. & T.J.; Funding Acquisition, T.J. & R.Y.

**Acknowledgments:** We thank Ekaterina Taralova for pioneering the project, Kui Tang for CRFs code, laboratory members for valuable comments and virus injections, Columbia Yeti shared High Performance Computing Cluster for computation resources, Stanford Neuroscience Gene Vector and Virus Core for AAVdj virus and the NEI (DP1EY024503, R01EY011787) for funding. This material is based upon work supported by the Defense Advanced Research Projects Agency (DARPA) under Contract No. HR0011-17-C-0026 and SIMPLEX N66001-15-C-4032 and in part by the U. S. Army Research Laboratory and the U. S. Army Research Office under contract number W911NF-12-1-0594 (MURI).. S.H. is a Howard Hughes Medical Institute International Student Research fellow. The authors declare no competing financial interests.

**Highlights**

- Decoding of visual stimuli from cortical responses using Conditional Random Field (CRF) models

- CRFs infer core cortical ensembles and network reconfiguration

- Optogenetics stimulation of high-ranked neurons trigger pattern completion

- CRFs capture network reconfiguration after optogenetic stimulation

**Summary**

A fundamental problem for machine learning in applications including natural language processing, computer vision and bioinformatics is the prediction of system states from measurements of variables with mutual dependencies. One solution is structured prediction methods that combine graphical models and classification algorithms, such as Conditional Random Fields (CRFs). However, the application of CRFs to infer the network structure and properties of neuronal groups with coordinated activity remains unexplored. We used CRFs on population two-photon calcium imaging data from primary visual cortex (V1) from awake head-fixed mice to identify cortical ensembles and predict visual stimuli. Using simultaneous two-photon optogenetics we show that high-ranked neurons identified with CRFs have pattern completion capability. Our method can also capture the reconfiguration in network properties of targeted neurons with single cell resolution. CRFs models could be broadly used for deciphering functional connectivity of neural circuits.

**Introduction**

The coordinated firing of neuronal populations is considered to be the substrate of sensory, behavioral and cognitive functions. As originally predicted by Hebb, coactive neuronal groups, defined as neuronal ensembles, are assumed to generate complex circuit functions ([Buzsaki, 2010](#_ENREF_9); [Cossart et al., 2003](#_ENREF_16); [Hebb, 1949](#_ENREF_25); [Luczak et al., 2009](#_ENREF_34); [Luczak et al., 2007](#_ENREF_35); [Mao et al., 2001](#_ENREF_36); [Miller et al., 2014](#_ENREF_38)) (Hebb, Buszaki, Harris, Mao, Cossart, Miller refs). Indeed, recent advances in two-photon calcium imaging and two-photon optogenetics have made possible the imprinting and recalling of cortical ensembles with single cell resolution in awake animals ([Carrillo-Reid et al., 2016](#_ENREF_13)). However, how the activation of specific groups of neurons relates to the network properties of cortical microcircuits has been difficult to elucidate. This is partly because it requires the online identification of single cells that can be targeted during close-loop optogenetic experiments, potentially under interventional manipulation of learned behavioral tasks.

One suitable method for online identification of ensembles is graph theory. Indeed, graph theory has been applied in neuroscience to model the structural and functional organization of the brain ([Bullmore and Sporns, 2009](#_ENREF_8)). However, in those models, graphs are usually constructed with nodes representing brain regions ([He et al., 2007](#_ENREF_24)), and edges representing information flow ([Iturria-Medina et al., 2008](#_ENREF_28)). For functional analysis, many studies have constructed graphs with data from fMRI, EEG and electrode arrays, taking brain regions ([Achard and Bullmore, 2007](#_ENREF_1); [Fair et al., 2008](#_ENREF_20); [Hagmann et al., 2008](#_ENREF_22)), voxels ([Eguiluz et al., 2005](#_ENREF_19); [van den Heuvel et al., 2008](#_ENREF_60); [Zuo et al., 2012](#_ENREF_64)) or electrode position ([Downes et al., 2012](#_ENREF_17)) as nodes, and activity associations such as cross correlation, mutual information and Granger causality as edges ([Bullmore and Sporns, 2009](#_ENREF_8); [Fair et al., 2008](#_ENREF_20); [Khazaee et al., 2015](#_ENREF_29); [Micheloyannis et al., 2009](#_ENREF_37); [Wang et al., 2010](#_ENREF_61)).

In addition, at the single cell level, graphical models have been used to describe organizing principles of artificial neural networks, identifying neurons that could have a potential role orchestrating the overall network activity ([Iturria-Medina et al., 2008](#_ENREF_28); [Sporns, 2000](#_ENREF_53)). Such graphs are usually associated with a restricted set of parameters that describe the weight and direction of edges obtained by pairwise correlations, therefore are limiting for characterizing the network structure and properties underlying the population activity. Finally, a few studies have applied graph theory to model network organization in calcium imaging data with single cell resolution in cultures or brain slices ([Bonifazi et al., 2009](#_ENREF_5); [Gururangan et al., 2014](#_ENREF_21); [Yatsenko et al., 2015](#_ENREF_62)), but these methods have not been applied to define the optimal configuration of neuronal ensembles that allows the prediction of external stimuli in awake animals.

Cortical ensembles in primary visual cortex consist of strongly interconnected neurons ([Carrillo-Reid et al., 2016](#_ENREF_13); [Cossart et al., 2003](#_ENREF_16); [Ko et al., 2011](#_ENREF_30); [Mao et al., 2001](#_ENREF_36); [Miller et al., 2014](#_ENREF_38)), forming a network structure that can be intuitively modeled with graph theory, where nodes and edges are biologically meaningful, representing neurons and their connections respectively. Here, we apply graph theory to the analysis of functional imaging data from two-photon calcium imaging of mouse visual cortex. We demonstrate that Conditional Random Field (CRF) models allow the identification of cortical ensembles associated with different experimental and physiological conditions, decoding the visual stimuli and predicting the neurons that are most efficient at pattern completion. This method opens the possibility of targeting with single cell resolution the most significant neurons from specific populations during microcircuit function.

**Results**

**CRF models of visual cortex population responses with two-photon calcium imaging**

CRFs model the conditional distribution *p*(**y**|**x**) of a network, where **x** represents observations and **y** represents true labels associated with a graphical structure ([Sutton and McCallum, 2012](#_ENREF_58)). Since no assumptions are made on **x**, CRFs can accurately describe the conditional distribution with complex dependencies in observation variables associated with a graphical structure that is used to constrain the interdependencies between labels. Therefore, CRFs have been successfully applied in diverse areas of machine learning such as analysis of texts ([Peng et al., 2011](#_ENREF_43)), bioinformatics ([Li et al., 2008](#_ENREF_32); [Liu et al., 2006](#_ENREF_33); [Sato and Sakakibara, 2005](#_ENREF_47)), computer vision ([He et al., 2004](#_ENREF_23); [Sminchisescu et al., 2006](#_ENREF_52)) and natural language processing ([Choi et al., 2005](#_ENREF_15); [Lafferty et al., 2001](#_ENREF_31)).

In order to study the network properties and structure of cortical ensembles we constructed CRF models using population responses to visual stimuli from layer 2/3 neurons of primary visual cortex in awake head-fixed mice (Figure 1A). We used high contrast oriented grating as visual stimuli. Population vectors representing the coordinated activity of neuronal groups were inferred from calcium imaging recordings ([Carrillo-Reid et al., 2015b](#_ENREF_11)) and used as training data (Figure 1B). We defined activity events from each neuron as nodes in an undirected graph, where each node can have two values: ‘0’ corresponding to non-activity, and ‘1’ corresponding to neuronal activity. In this way nodes interact with each other by connecting edges, which have four possible combinations ‘00’, ‘01’, ‘10’, and ‘11’, depending on the values of the two nodes on the edge. The two values associated with nodes and the four values associated with edges are characterized by a set of parameters called node potentials and edge potentials correspondingly (Figure 1C). These parameters are also known as potential functions and reflect the scores of individual values on each node and edge. Using part of the observation data, we first estimated model parameters and then performed cross-validation on held-out data (see Experimental Procedures). Need numbers and statistics. The final normalized product of the corresponding nodes and edge potentials describes the likelihood that a given neuronal population exhibits a specific activation pattern. This could also be understood as the probability to estimate different network states from observed population vectors (Figure S1).

To integrate information of the external stimulus along with the observed neuronal data, we added an additional node for each type of stimulus that was presented to the animal. This node was set to ‘1’ when the corresponding stimulus was on and ‘0’ when the stimulus was off (Figures 1C and 1D). The general and mathematical properties of CRF models obtained with added nodes did not significantly differ from CRF models obtained without added nodes (Figure S2). In both conditions, CRFs modeled the conditional probability of network states given the observations. Therefore, by treating visual stimuli as added nodes and comparing the output likelihood of observing each stimulus, CRFs were able to predict visual stimuli from observed data. In this way, the nodes directly connected to the added nodes represent a specific model for different visual stimuli (Figure 1D). Given two different visual stimuli (horizontal or vertical drifting gratings), the likelihood corresponding to observing each stimulus is defined by , and . Thus, the relative likelihood can be used to classify the presented stimuli (Figure 1E). To evaluate the classification performance for each model, we examined the area under the curve (AUC) from the receiver operating characteristic (ROC) curve for the two different stimuli (AUC, mean ± SEM: 0.8319 ± 0.0184 for horizontal, 0.8455 ± 0.0098 for vertical). These results demonstrated that CRFs are able to predict with similar performance different orientations of drifting-gratings from calcium imaging population recordings.

**Identification of core neurons in cortical ensembles with CRFs**

Coactive cortical ensembles represent neuronal populations with modular properties ([Carrillo-Reid et al., 2015a](#_ENREF_10); [Carrillo-Reid et al., 2016](#_ENREF_13); [Cossart et al.](#_ENREF_16); [Mao et al.](#_ENREF_36); [Miller et al., 2014](#_ENREF_38)). Structural and functional modularization, in both macro-scale and micro-scale levels, are characterized by local structures with high inter-connectivity, where a group of neurons shows dense physical or functional connections ([Achard et al., 2006](#_ENREF_2); [Bonifazi et al., 2009](#_ENREF_5); [Hagmann et al., 2008](#_ENREF_22); [He et al., 2007](#_ENREF_24); [Shimono and Beggs, 2015](#_ENREF_50); [Sporns et al., 2007](#_ENREF_54); [Stetter et al., 2012](#_ENREF_55); [Zuo et al., 2012](#_ENREF_64)). Such structures can be described by different concepts such as cliques, communities ([Palla et al., 2005](#_ENREF_42)), hubs and modules ([Bullmore and Sporns, 2009](#_ENREF_8)).

Given the intrinsic variability in neural responses, in order to design close-loop optogenetic experiments with single cell resolution, it is necessary to identify these cortical modules formed by “core neurons” from cortical ensembles that can efficiently represent different visual stimuli ([Miller et al., 2014](#_ENREF_38); [Sadovsky and MacLean, 2014](#_ENREF_45)). The classification nature of CRFs provides a convenient way to define core neurons from cortical ensembles. To identify core ensembles we set the activity of each neuron to be either ‘1’ or ‘0’ in all population activity vectors of the dataset, and compared the output likelihood using the inferred CRF models (Figure 2A). Then, we calculated single neuron preference by binarizing the likelihood difference (Figure 2B). Since core neurons are likely to have concomitant activity, we defined the node strength in CRFs models as the summation of edge potentials ( terms) from all connecting edges for each node in the graph. In this way, highly connected neurons or strongly connecting neurons could have high node strength whereas weakly connected neurons have low node strength (Figure 2C). We defined core neurons from cortical ensembles as the neurons that can be used to predict each visual stimulus with higher performance (AUC) and have high node strength (Figure 2D). To demonstrate the general applicability of our method, we analyzed publically open datasets (Allen Brain Observatory) that contains data from layer 2/3 of primary visual cortex consisting in several visual stimuli types with different experimental settings, and showed that our approach is able to find the core ensembles for each visual stimulus (Figures 3; overall classification AUC for TF = 1: 0.8514 ± 0.0429 [mean ± S.D.]). Need numbers and statistics. Interestingly, our results demonstrated that ensemble classification performance to different drifting-gratings is significantly better for low temporal frequencies as measured by ROC curves and AUC values (Figures 3C-3E; classification AUC [mean ± S.E.M]: AUCTF=1 = 0.8514 ± 0.0096, AUCTF=2 = 0.8409 ± 0.0097, AUCTF=4 = 0.8007 ± 0.0117, AUCTF=8 = 0.7252 ± 0.0165, AUCTF=15 = 0.6554 ± 0.0124; p1,4<0.01, p1,8, p1,15<0.001; n = 5 animals, 20 ensembles; Wilcoxon rank sum test). Need numbers and statistics.

**Prediction performance of core neurons**

We next investigated whether these core neurons from cortical ensembles have the better performance for the prediction of visual stimuli. To do so, we randomly resized the population vectors containing the core neurons from cortical ensembles by adding or removing elements from the group, and examined the stimulus prediction performance. The similarity function and prediction performance of population vectors formed with the core neurons from cortical ensembles has a maximum value when the size of population vectors is unchanged (Figures 4A-4C; original core neurons similarity [mean ± S.D.] 0.2887 ± 0.0926; prediction AUC 0.9383 ± 0.0333). Need numbers and statistics. We also calculated three standard measurements from the number of true positives (TP), true negatives (TN), false positives (FP) and false negatives (FN): accuracy, defined as (TP+TN)/(TP+TN+FP+FN); precision, defined as TP/(TP+FP); and recall, defined as TP/(TP+FN). Using these measurements, we demonstrated that population vectors from core ensembles achieve the best accuracy, precision and recall when predicting the presented visual stimuli, compared with resized ensembles (Figure S3; original core neurons prediction accuracy 0.8367 ± 0.0623, precision 0.6175 ± 0.1752, recall 0.9067 ± 0.0717). Need numbers and statistics.

These results showed that population vectors formed with the core neurons from cortical ensembles identified by CRFs represent an efficient population to predict external visual stimuli. This fact raises the question of whether such population vectors are a specific non-random subgroup. To answer this question, we randomly sampled a subset from all the population of neurons ranging from 10% to 100% of the total size of core ensembles. We observed that the prediction performance from random groups of neurons was significantly lower than the core ensembles (Figures 4D-4F and S3; best similarity [mean ± S.D.] 0.1985 ± 0.0590, AUC 0.5029 ± 0.0616, accuracy 0.6906 ± 0.0671, precision 0.2570 ± 0.1756, recall 0.3617 ± 0.1027), Need numbers and statistics. indicating that the core neurons from cortical ensembles achieve the best classification performance and cannot be considered as random neuronal groups.

**Comparison with previously used approaches**

The population vectors formed by the core neurons from cortical ensembles identified with CRFs have comparable prediction performance (AUC [mean ± S.E.M] 0.9383 ± 0.0074, accuracy 0.8367 ± 0.0155, precision 0.6175 ± 0.0402, recall 0.9067 ± 0.0165; n = 6 mice, 20 ensembles) with previously used methods for cortical ensemble classification using SVD ([Carrillo-Reid et al., 2015a](#_ENREF_10); [Carrillo-Reid et al., 2015b](#_ENREF_11); [Carrillo-Reid et al., 2016](#_ENREF_13)) (Figures 5A-5E; AUC 0.9111 ± 0.0103, p>0.05; accuracy 0.8448 ± 0.0157, p>0.05; precision 0.6724 ± 0.0446, p>0.05; recall 0.8097 ± 0.0245, p<0.01; n = 6 mice, 19 ensembles) or groups of neurons with high orientation selectivity index (OSI) (Figures S4A-S4E; AUC 0.9341 ± 0.0087, p>0.05; accuracy 0.9217 ± 0.0136, p<0.001; precision 0.8353 ± 0.0343, p<0.001; recall 0.8650 ± 0.0294, p>0.05; n = 6 mice, 20 ensembles, Wilcoxon rank sum test). Need numbers and statistics. The core neurons identified with CRFs shared 51.29% ± 5.51% (S.E.M) cells with the SVD ensembles (Figure 5F), and were composed of a mixed population of cells with high OSI (42.34% ± 3.84% [S.E.M.]) and cells with low OSI (Figure S4F), indicating that the core neurons from cortical ensembles identified with CRFs are not simply orientation selective cells. Need numbers and statistics. The percentage from the whole population of neurons from SVD core ensembles (Figures 5G and 5H) and high OSI ensembles (Figures S4G and S4H) was not significantly different from population vectors defined by CRFs core ensembles ([mean ± S.E.M] CRF: 11.43% ± 0.95%; SVD: 10.49% ± 1.35%, p>0.05; OSI: 12.65% ± 2.26%, p>0.05; n = 6 mice, 20 ensembles, Wilcoxon rank sum test). Need numbers and statistics. Since classification performance of CRFs is similar to previously used methods for cortical ensemble identification the main advantage of CRFs is their ability to construct models that capture network properties that could be used to target individual neurons.

**Identification of pattern completion neurons using CRFs models**

The repetitive activation of an identified neuronal population with two-photon optogenetics imprints an artificial cortical ensemble that can be recalled later on by specific members of the ensemble ([Carrillo-Reid et al., 2016](#_ENREF_13)). Since CRFs can be used to identify the core neurons from cortical ensembles, we hypothesize that the same approach could also be used for the identification of neurons with pattern completion capability from these artificially imprinted cortical ensembles.

We used the structural and performance parameters from CRFs applied to simultaneous two-photon imaging and two-photon optogenetic experiments with single cell resolution (Figure S5A) to define high-ranked neurons as the ones with strong node strength and high AUC values (Figure S5B). Our analytical approach demonstrated that single-cell two-photon optogenetic stimulation of high-ranked neurons (AUC 0.8397 ± 0.0361 [mean ± S.E.M]; node strength -0.1405 ± 0.0770; n = 3 neurons) was able to evoke pattern completion of artificially imprinted ensembles, whereas non-high-ranked neurons (AUC 0.5680 ± 0.0292; node strength -1.0332 ± 0.0573; n = 5 neurons) were unable to recall imprinted cortical ensembles (Figures S5C and S5D). Need numbers and statistics. These experiments support the hypothesis that neurons with pattern completion capability represent neurons highly connected with other members of artificially imprinted cortical ensembles ([Carrillo-Reid et al., 2016](#_ENREF_13)).

**Reconfiguration of cortical microcircuits using CRFs models**

To investigate if our approach can also describe changes in network connectivity under different experimental conditions, we compared the models generated by CRFs before and after two-photon population manipulation of a given set of neurons for several times (Figure 6A), an experimental protocol that reconfigures network activity building new coactive ensembles ([Carrillo-Reid et al., 2016](#_ENREF_13)). To visualize the change in graph connectivity induced by the imprinting protocol in neurons with pattern completion capability we constructed isomorphic graphs from the CRFs models and arrange them in a circular configuration (Figure 6B). After the artificial ensemble is imprinted, neurons with pattern completion capability show better predictive performance and higher node strength (Figure 6C; node strength increased from -0.1539 to 0.1609, prediction AUC increased from 0.5234 to 0.7447; n = 1 neuron), Need numbers and statistics. This demonstrated that structural and prediction parameters inferred from CRFs can be used to study changes in network properties of specific neurons and that CRFs could be used to target single neurons that play a key role in the computational properties of cortical microcircuits. Interestingly, while the node degree and centrality of artificially stimulated neurons was significantly increased (Figures 6D and 6E; node degree from 0.0489 ± 0.0048 [mean ± S.E.M] to 0.0689 ± 0.0033, p<0.01; centrality from 0.3847 ± 0.0428 to 0.6643 ± 0.0482, p<0.001), the node strength and clustering coefficient from the photostimulated population remained stable (Figures 6F and 6G; node strength from -0.2427 ± 0.0499 to -0.1528 ± 0.0564, p>0.05; clustering coefficient from 0.2844 ± 0.0324 to 0.2490 ± 0.0156, p>0.05; n = 15 neurons; Wilcoxon signed rank test). Need numbers and statistics. On the other hand, for non-photostimulated neurons the graphical properties of CRFs before and after population photostimulation remained stable (Figure S6) suggesting that imprinted ensembles have been added to cortical microcircuits but preserving a balance with the overall network structure.

The fact that CRFs were able to describe changes in the reconfiguration of network connectivity from artificially imprinted ensembles demonstrates the potential of structured prediction methods to study the modulation of neuronal microcircuits induced by external perturbations or pathological conditions.

**Discussion**

**Machine learning analysis of functional connectivity in cortical microcircuits**

In this study, we provide a tool for modeling network properties of mouse primary visual cortex *in vivo* using Conditional Random Fields (CRFs), a novel machine learning graphical method. As opposite to traditional static descriptions of the data, structured prediction methods such as CFRs not only reveal the structure of the functional connections, but also provide the conditional probability of the interactions between neurons to find neuronal ensembles capable of predicting specific sensory stimuli.

In the past decades, graph theory has been applied to characterize the structure and function of neuronal networks ([Achard and Bullmore, 2007](#_ENREF_1); [Bettencourt et al., 2007](#_ENREF_4); [Chiang et al., 2016](#_ENREF_14); [Downes et al., 2012](#_ENREF_17); [Fair et al., 2008](#_ENREF_20); [Hagmann et al., 2008](#_ENREF_22); [Iturria-Medina et al., 2008](#_ENREF_28); [Oh et al., 2014](#_ENREF_40); [Supekar et al., 2008](#_ENREF_57); [Yu et al., 2008](#_ENREF_63); [Zuo et al., 2012](#_ENREF_64)). While most of these studies operated on functional recordings across multiple brain regions ([Achard and Bullmore, 2007](#_ENREF_1); [Chiang et al., 2016](#_ENREF_14); [Fair et al., 2008](#_ENREF_20); [Hinne et al., 2013](#_ENREF_26); [Zuo et al., 2012](#_ENREF_64)), only a few have focused on the general network properties of cortical circuits with recordings from single neurons ([Bonifazi et al., 2009](#_ENREF_5); [Sadovsky and MacLean, 2014](#_ENREF_45); [Stetter et al., 2012](#_ENREF_55); [Yatsenko et al., 2015](#_ENREF_62)).

The majority of methods applied to infer network properties in brain slices ([Cossart et al., 2003](#_ENREF_16); [Ikegaya et al., 2004](#_ENREF_27); [Mao et al., 2001](#_ENREF_36); [Sadovsky and MacLean, 2014](#_ENREF_45); [Stetter et al., 2012](#_ENREF_55)) or *in vivo* ([Yatsenko et al., 2015](#_ENREF_62)) operate on the correlation matrix, and aim to recover the functional dependencies between observed neurons. Such methods are valuable for revealing some properties such as node degrees, clustering coefficients or functional hubs. However, these methods are model-free, therefore are incapable of describing the overall network dynamics based on the probability distribution of neuronal ensembles. Our method provides an alternative by directly modeling the statistical dependencies of each neuron.

**CRFs graphical models identify neuronal ensembles**

Compared with generative models that make assumptions on the dependencies between all the observed variables from the model, CRFs only model the hidden system states dependent on observed features. Since no independence assumptions are made between observed variables, CRFs avoid potential errors introduced by unobserved common inputs. Additionally, given the finite number of network states described by population activity, the conditional distribution is sufficient for making predictions, both for the population state and for identifying core neurons in each state. Another important difference between CRFs and generative models is that the ability to model arbitrary dependencies between observed variables is restricted in generative models. Compared with other discriminative finite-state models such as Maximum Entropy Markov Models (MEMM), CRFs use global normalizers to overcome the local bias in MEMM induced by local normalizers, and have been shown to achieve higher accuracy in diverse applications ([Lafferty et al., 2001](#_ENREF_31)). Therefore, CRFs appear to be promising for modeling cortical functional connectivity and for identifying a core cortical ensemble that could be easily manipulated by two-photon optogenetics.

The difficulty of constructing CRFs lies in the computation of global normalizers. With an arbitrary graph structure, this problem is often intractable. Recent advances that combines Bethe free energy approximation and Frank-Wolfe methods for inference and learning model parameters allow fast and relatively accurate construction of cyclic CRFs ([Tang et al., 2016](#_ENREF_59)). Thus, CRFs can be applied to datasets with thousands of interconnected neurons. However, a main constraint for applying CRFs and any machine learning approach is the number of samples in the training dataset.

**Comparison with classification algorithms to detect cortical ensembles**

The overall activity of multiple cells at a given time window can be understood as a multidimensional array of population vectors where vectors pointing to a similar point in space can be considered as a group. We previously showed that population vectors defining a group (i.e. a cortical ensemble) can be extracted from multidimensional arrays by performing singular value decomposition (SVD) ([Carrillo-Reid et al., 2015b](#_ENREF_11)). Even though SVD can identify cortical ensembles reliably, it lacks a structured graphical model that allows the systematic study of changes in network properties.

Compared with the SVD method for neuronal ensemble identification ([Carrillo-Reid et al., 2015a](#_ENREF_10); [Carrillo-Reid et al., 2015b](#_ENREF_11)), our approach modestly improved prediction performance. One reason could be that current CRF learning algorithms separately perform the structure learning and parameter learning steps. Therefore, the learned graphical structure and parameters may not be the globally best matching ones. However, it is still computational unrealistic to explore all possible structures and parameter combinations. Additionally, approximations during the parameter learning step can sometimes compromise the global optimality guarantees.

**Identification of key neurons for optogenetic targeting**

Electrical stimulation of visual cortex has been used for decades as an attempt to provide useful visual sensations to patients that have lost the functionality of their eyes ([Brindley Gs Fau - Lewin and Lewin, 1986](#_ENREF_6)). The sensations produced by electrical stimulation of the visual cortex were termed phosphenes since they represented bright spots. A challenging issue regarding prostheses is the training of patients using devices with a large number of electrodes ([Shepherd et al., 2013](#_ENREF_49)). Our results suggest that after a given network have been imprinted ([Carrillo-Reid et al., 2016](#_ENREF_13)), the identification of neurons with pattern completion capability could be used to reduce the number of active points that require stimulation. The further development of network models based on population activity that can predict a given set of features embedded in visual stimuli will be crucial for the efficient manipulation of cortical ensembles.

It has been shown that the connectivity of diverse systems described by graphs with complex topologies follow a scale-free power-law distribution ([Barabasi and Albert, 1999](#_ENREF_3)). Scale-free networks are characterized by the existence of a small subset of nodes with high connectivity ([Carrillo-Reid et al., 2015a](#_ENREF_10)). Similarly, cortical ensembles described by CRFs could be characterized by a subset of neurons with strong synaptic connections. The existence of neurons with pattern completion capability has been demonstrated in previous studies where perturbing the activity of single neurons was able to change the overall network dynamics ([Bonifazi et al., 2009](#_ENREF_5); [Carrillo-Reid et al., 2016](#_ENREF_13); [Hagmann et al., 2008](#_ENREF_22)). However the efficient identification of such neurons from network models that capture the overall properties of neuronal microcircuits was lacking.

Our results suggest that the structural and predictive parameters defined by CRFs models could be used in the design of closed loop experiments with single cell resolution to investigate the role of a specific subpopulation of neurons in a given cortical microcircuit during different behavioral events.

**Figure legends**

**Figure 1. Classification of visual stimuli from calcium population responses using CRFs**

(A) Experimental setup: simultaneous two-photon imaging and two-photon optogenetics were performed in layer 2/3 of primary visual cortex in head fixed freely moving mice. (B) Neurons were automatically detected from each frame and multidimensional population vectors representing coactive groups of neurons were defined from binary data (Related to Figure S1). (C) Graphical representation of CRFs. Circles represent neurons. Squares represent added nodes depicting visual stimuli. Shaded nodes (x) represent observed data. White nodes (y) represent true states of the neurons, and are connected by edges that indicate their mutual dependencies. node potentials are defined over the two possible states of each node, and edge potentials are defined over the four possible states of each existing edge, depending on the state of the two nodes it connects. (D) Example of a CRF graph constructed from real data. In this case, added nodes (squares) represent horizontal (red) and vertical (blue) drifting-gratings. First- and second-degree connections with added nodes are highlighted in the corresponding color. Scale bar: 50μm. (E) Relative likelihood predicting horizontal (red; top) and vertical (blue; bottom) drifting-gratings calculated by CRFs. Colored stripes indicate visual stimuli. Scale bar: 10 seconds.

**Figure 2. Identification of core neurons from cortical ensembles with CRFs**

(A) Schematic representation of ensemble identification from CRF models with added nodes representing different visual stimuli (Related to Figure S2). The activity of the neuron is set to ‘1’ or ‘0’ at each frame, and the likelihood and of modified population vectors is calculated. Edge color tone represents edge potential strength (); node color represents node strength. Scale bar: 50μm. (B) Likelihood inference and prediction for neurons belonging to cortical ensembles representing horizontal (top; red) or vertical (bottom; blue) drifting-gratings. (C) Graphical representation of node strength magnitude. (D) Core neurons from cortical ensembles for two visual stimuli defined as neurons with high AUC and high node strength (top right quadrant). Confidence levels were defined from CRF models of shuffled data (grey bars).

**Figure 3. Cortical ensembles have better performance at low temporal frequencies**

(A) Graphical representation of CRFs from the Allen Brain Observatory dataset, with four orientations of drifting-gratings. Edge color indicates the strength of inferred connections; node size indicates node degree. (B) Temporal course of ensemble classification using CRFs. Colored stripes indicate different visual stimuli. Scale bar: 200 frames. (C) ROC curves of core neurons from cortical ensembles for different temporal frequencies (TF: 1, 2, 4, 8 and 15 Hz) (D) AUC for different TFs compared to TF=1 Hz (classification AUC: AUCTF=1 = 0.8514 ± 0.0430, AUCTF=2 = 0.8409 ± 0.0433 [S.D.], AUCTF=4 = 0.8007 ± 0.0524, AUCTF=8 = 0.7252 ± 0.0737, AUCTF=15 = 0.6554 ± 0.0556; p1,2=0.4735 n.s; p1,4=0.0036\*\*; p1,8=2.062e-06\*\*\*; p1,15=6.7956e-08\*\*\*) . Note that cortical ensembles have better prediction performance for low TFs. Dashed line represents random classification performance. (E) Preferred orientation selectivity of core cortical ensembles for TF=1Hz. The radius of each circle depicts AUC values from zero (center) to 1 (border). Dotted inner circles represent random performance (AUC=0.5). Data presented as box and whisker plots displaying median and interquartile ranges (n = 5 mice, 20 ensembles; Wilcoxon rank sum test).

**Figure 4. Core neurons from cortical ensembles have the highest classification performance for visual stimuli**

(A) Cosine similarity and between core population vectors identified with CRFs and randomly down-sampled or up-sampled population vectors. The cosine similarity of population vectors belonging to non-predicted visual stimuli is shown in grey. (B) AUC values from core population vectors and (C) ROC curves from randomly down-sampled or up-sampled population vectors. Note that randomly removing or adding elements from the core ensemble decreases the ability to predict visual stimuli (Related to Figure S3). (D) Cosine similarity, (E) AUC values and (F) ROC curves from randomly chosen core ensembles of different sizes. Data presented as box and whisker plots displaying median and interquartile ranges (n=6 mice, 20 ensembles; Wilcoxon rank sum test).

**Figure 5. Comparison of classification performance between CRF and SVD core ensembles**

(A) ROC curves from core ensembles identified with CRFs (orange) or SVD (green). Dashed line represents random classification performance. (B) AUC values for core ensembles identified with CRFs (orange) or SVD (green) (P=0.0657 n.s). (C) Accuracy (P=0.5458 n.s), (D) precision (P=0.3915 n.s) and (E) recall (P=0.0028\*\*) of classification performance for different visual stimuli using CRFs (orange) or SVD (green). (F) Percentage of shared neurons between CRF and SVD core ensembles. (G) Percentage from the total population size representing core ensembles from CRFs and SVD (P=0.0865 n.s). (H) Spatial maps of core ensembles from CRF (orange) and SVD (green) for two different orientations of drifting-gratings. Neurons shared between CRF and SVD core ensembles are represented by orange dots circled by green. Scale bar: 50μm. Data presented as box and whisker plots displaying median and interquartile ranges (n=6 mice; Wilcoxon rank sum test). Related to Figure S4.

**Figure 6. Reconfiguration of cortical microcircuits by two-photon optogenetic stimulation**

(A) Graphical models obtained using CRFs from simultaneous two-photon imaging and two-photon optogenetic stimulation of a neuron with pattern completion capability (Related to Figure S5) before (left) and after (right) two-photon optogenetic ensemble imprinting. Square on bottom left represents added node for optogenetic stimuli. Edge color tone represents edge potential strength (); node color represents node strength. Node size represents node degree. Scale bar: 50μm. (B) Isomorphic graphs of CRFs models before (pre) and after (post) ensemble imprinting. Blue neurons were stimulated with two-photon optogenetics (60 trials; 4Hz). Connections between photostimulated neurons are shown in blue. Red dot represents stimulated neuron. (C) Node strength and AUC values showed changes in network reconfiguration of neurons with pattern completion capability. Stimulated neuron is represented in red before (left) and after (right) ensemble imprinting. Confidence levels calculated from random data are depicted by grey bars. (D) Node degree (P=0.0015\*\*), (E) centrality (P=0.0002\*\*\*), (F) node strength (P=0.1876 n.s) and (G) clustering coefficient (P=0.3258 n.s) of neurons belonging to imprinted ensemble before (pre) and after (post) photostimulation protocol. Data plots represent photostimulated neurons activated by imprinting protocol. Black lines highlight neurons with pattern completion capability. Note that measurements related to functional connectivity were significantly increased by imprinting protocol whereas values related to the contribution of photostimulated neurons into the whole network remain stable indicating reconfiguration of a subpopulation of neurons but preserving global network properties (n=1 mouse, 15 neurons; Wilcoxon signed rank test). Related to Figure S6.

**References**

Achard, S., and Bullmore, E. (2007). Efficiency and cost of economical brain functional networks. PLoS computational biology *3*, e17.

Achard, S., Salvador, R., Whitcher, B., Suckling, J., and Bullmore, E. (2006). A resilient, low-frequency, small-world human brain functional network with highly connected association cortical hubs. J Neurosci *26*, 63-72.

Barabasi, A.L., and Albert, R. (1999). Emergence of scaling in random networks. Science (New York, NY) *286*, 509-512.

Bettencourt, L.M., Stephens, G.J., Ham, M.I., and Gross, G.W. (2007). Functional structure of cortical neuronal networks grown in vitro. Physical review E, Statistical, nonlinear, and soft matter physics *75*, 021915.

Bonifazi, P., Goldin, M., Picardo, M.A., Jorquera, I., Cattani, A., Bianconi, G., Represa, A., Ben-Ari, Y., and Cossart, R. (2009). GABAergic hub neurons orchestrate synchrony in developing hippocampal networks. Science (New York, NY) *326*, 1419-1424.

Brindley Gs Fau - Lewin, W.S., and Lewin, W.S. (1986). The sensations produced by electrical stimulation of the visual cortex.

Brown, S.L., Joseph, J., and Stopfer, M. (2005). Encoding a temporally structured stimulus with a temporally structured neural representation. Nature neuroscience *8*, 1568-1576.

Bullmore, E., and Sporns, O. (2009). Complex brain networks: graph theoretical analysis of structural and functional systems. Nature reviews Neuroscience *10*, 186-198.

Buzsaki, G. (2010). Neural syntax: cell assemblies, synapsembles, and readers. Neuron *63*.

Carrillo-Reid, L., Lopez-Huerta, V.G., Garcia-Munoz, M., Theiss, S., and Arbuthnott, G.W. (2015a). Cell Assembly Signatures Defined by Short-Term Synaptic Plasticity in Cortical Networks. International journal of neural systems *25*, 1550026.

Carrillo-Reid, L., Miller, J.E., Hamm, J.P., Jackson, J., and Yuste, R. (2015b). Endogenous sequential cortical activity evoked by visual stimuli. J Neurosci *35*, 8813-8828.

Carrillo-Reid, L., Tecuapetla, F., Tapia, D., Hernandez-Cruz, A., Galarraga, E., Drucker-Colin, R., and Bargas, J. (2008). Encoding network states by striatal cell assemblies. J Neurophysiol *99*, 1435-1450.

Carrillo-Reid, L., Yang, W., Bando, Y., Peterka, D.S., and Yuste, R. (2016). Imprinting and recalling cortical ensembles. Science (New York, NY) *353*, 691-694.

Chiang, S., Cassese, A., Guindani, M., Vannucci, M., Yeh, H.J., Haneef, Z., and Stern, J.M. (2016). Time-dependence of graph theory metrics in functional connectivity analysis. NeuroImage *125*, 601-615.

Choi, Y., Cardie, C., Riloff, E., and Patwardhan, S. (2005). Identifying sources of opinions with conditional random fields and extraction patterns. Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing HLT 05, 355-362.

Cossart, R., Aronov, D., and Yuste, R. (2003). Attractor dynamics of network UP states in the neocortex. Nature *423*, 283-288.

Downes, J.H., Hammond, M.W., Xydas, D., Spencer, M.C., Becerra, V.M., Warwick, K., Whalley, B.J., and Nasuto, S.J. (2012). Emergence of a small-world functional network in cultured neurons. PLoS computational biology *8*, e1002522.

Drew, P.J., Shih, A.Y., Driscoll, J.D., Knutsen, P.M., Blinder, P., Davalos, D., Akassoglou, K., Tsai, P.S., and Kleinfeld, D. (2010). Chronic optical access through a polished and reinforced thinned skull. Nature methods *7*, 981-984.

Eguiluz, V.M., Chialvo, D.R., Cecchi, G.A., Baliki, M., and Apkarian, A.V. (2005). Scale-free brain functional networks. Physical review letters *94*, 018102.

Fair, D.A., Cohen, A.L., Dosenbach, N.U., Church, J.A., Miezin, F.M., Barch, D.M., Raichle, M.E., Petersen, S.E., and Schlaggar, B.L. (2008). The maturing architecture of the brain's default network. Proceedings of the National Academy of Sciences of the United States of America *105*, 4028-4032.

Gururangan, S.S., Sadovsky, A.J., and MacLean, J.N. (2014). Analysis of graph invariants in functional neocortical circuitry reveals generalized features common to three areas of sensory cortex. PLoS computational biology *10*, e1003710.

Hagmann, P., Cammoun, L., Gigandet, X., Meuli, R., Honey, C.J., Wedeen, V.J., and Sporns, O. (2008). Mapping the structural core of human cerebral cortex. PLoS biology *6*, e159.

He, X., Zemel, R.S., and Carreira-Perpinan, M.A. (2004). Multiscale conditional random fields for image labeling. Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition *2*, 695 -702.

He, Y., Chen, Z.J., and Evans, A.C. (2007). Small-world anatomical networks in the human brain revealed by cortical thickness from MRI. Cerebral cortex (New York, NY : 1991) *17*, 2407-2419.

Hebb, D.O. (1949). The Organization of Behaviour (New York: Wiley).

Hinne, M., Heskes, T., Beckmann, C.F., and van Gerven, M.A.J. (2013). Bayesian inference of structural brain networks. NeuroImage *66*, 543-552.

Ikegaya, Y., Aaron, G., Cossart, R., Aronov, D., Lampl, I., Ferster, D., and Yuste, R. (2004). Synfire chains and cortical songs: temporal modules of cortical activity. Science (New York, NY) *304*, 559-564.

Iturria-Medina, Y., Sotero, R.C., Canales-Rodriguez, E.J., Aleman-Gomez, Y., and Melie-Garcia, L. (2008). Studying the human brain anatomical network via diffusion-weighted MRI and Graph Theory. NeuroImage *40*, 1064-1076.

Khazaee, A., Ebrahimzadeh, A., and Babajani-Feremi, A. (2015). Identifying patients with Alzheimer's disease using resting-state fMRI and graph theory. Clinical neurophysiology : official journal of the International Federation of Clinical Neurophysiology *126*, 2132-2141.

Ko, H., Hofer, S.B., Pichler, B., Buchanan, K.A., Sjostrom, P.J., and Mrsic-Flogel, T.D. (2011). Functional specificity of local synaptic connections in neocortical networks. Nature *473*, 87-91.

Lafferty, J., McCallum, A., and Pereira, F.C.N. (2001). Conditional random fields: Probabilistic models for segmenting and labeling sequence data. ICML '01 Proceedings of the Eighteenth International Conference on Machine Learning *8*, 282-289.

Li, C.T., Yuan, Y., and Wilson, R. (2008). An unsupervised conditional random fields approach for clustering gene expression time series. Bioinformatics (Oxford, England) *24*, 2467-2473.

Liu, Y., Carbonell, J., Weigele, P., and Gopalakrishnan, V. (2006). Protein fold recognition using segmentation conditional random fields (SCRFs). Journal of computational biology : a journal of computational molecular cell biology *13*, 394-406.

Luczak, A., P., B., and Harris, K.D. (2009). Spontaneous events outline the realm of possible sensory responses in neocortical populations. Neuron *62*.

Luczak, A., P., B., L., M.S., G., B., and Harris, K.D. (2007). Sequential structure of neocortical spontaneous activity in vivo. Proc Natl Acad Sci U S A *104*.

Mao, B.Q., F., H.-S., Aronov, D., Froemke, R.C., and Yuste, R. (2001). Dynamics of spontaneous activity in neocortical slices. Neuron *32*, 833-898.

Micheloyannis, S., Vourkas, M., Tsirka, V., Karakonstantaki, E., Kanatsouli, K., and Stam, C.J. (2009). The influence of ageing on complex brain networks: a graph theoretical analysis. Human brain mapping *30*, 200-208.

Miller, J.E., Ayzenshtat, I., Carrillo-Reid, L., and Yuste, R. (2014). Visual stimuli recruit intrinsically generated cortical ensembles. Proceedings of the National Academy of Sciences of the United States of America *111*, E4053-4061.

Mukamel, E.A., Nimmerjahn, A., and Schnitzer, M.J. (2009). Automated analysis of cellular signals from large-scale calcium imaging data. Neuron *63*, 747-760.

Oh, S.W., Harris, J.A., Ng, L., Winslow, B., Cain, N., Mihalas, S., Wang, Q., Lau, C., Kuan, L., Henry, A.M.*, et al.* (2014). A mesoscale connectome of the mouse brain. Nature *508*, 207-214.

Packer, A.M., Peterka, D.S., Hirtz, J.J., Prakash, R., Deisseroth, K., and Yuste, R. (2012). Two-photon optogenetics of dendritic spines and neural circuits. Nature methods *9*, 1202-1205.

Palla, G., Derényi, I., Farkas, I., and Vicsek, T. (2005). Uncovering the overlapping community structure of complex networks in nature and society. Nature *435*, 814-818.

Peng, H.-K., Zhu, J., Piao, D., Yan, R., and Zhang, Y. (2011). Retweet Modeling Using Conditional Random Fields. In 2011 IEEE 11th International Conference on Data Mining Workshops (IEEE), pp. 336-343.

Ravikumar, P., Wainwright, M.J., and Lafferty, J.D. (2010). High-dimensional Ising model selection using ℓ1-regularized logistic regression. The Annals of Statistics *38*, 1287-1319.

Sadovsky, A.J., and MacLean, J.N. (2014). Mouse visual neocortex supports multiple stereotyped patterns of microcircuit activity. J Neurosci *34*, 7769-7777.

Sasaki, T., Matsuki, N., and Ikegaya, Y. (2007). Metastability of active CA3 networks. The Journal of neuroscience : the official journal of the Society for Neuroscience *27*, 517-528.

Sato, K., and Sakakibara, Y. (2005). RNA secondary structural alignment with conditional random fields. Bioinformatics (Oxford, England) *21 Suppl 2*, ii237-242.

Schreiber, S., Fellous, J.M., Whitmer, D., Tiesinga, P., and Sejnowski, T.J. (2003). A new correlation-based measure of spike timing reliability. Neurocomputing *52-4*, 925-931.

Shepherd, R.K., Shivdasani, M.F., Nayagam, D.F., Williams, C.F., and Blamey, P.J. (2013). Visual prostheses for the blind. Trends Biotechnol *10*.

Shimono, M., and Beggs, J.M. (2015). Functional Clusters, Hubs, and Communities in the Cortical Microconnectome. Cerebral cortex (New York, NY : 1991) *25*, 3743-3757.

Shmiel, T., Drori, R., Shmiel, O., Ben-Shaul, Y., Nadasdy, Z., Shemesh, M., Teicher, M., and Abeles, M. (2006). Temporally precise cortical firing patterns are associated with distinct action segments. J Neurophysiol *96*, 2645-2652.

Sminchisescu, C., Kanaujia, A., and Metaxas, D. (2006). Conditional models for contextual human motion recognition. Computer Vision and Image Understanding *104*, 210-220.

Sporns, O. (2000). Theoretical Neuroanatomy: Relating Anatomical and Functional Connectivity in Graphs and Cortical Connection Matrices. Cerebral Cortex *10*, 127-141.

Sporns, O., Honey, C.J., and Kotter, R. (2007). Identification and classification of hubs in brain networks. PloS one *2*, e1049.

Stetter, O., Battaglia, D., Soriano, J., and Geisel, T. (2012). Model-free reconstruction of excitatory neuronal connectivity from calcium imaging signals. PLoS computational biology *8*, e1002653.

Stopfer, M., Jayaraman, V., and Laurent, G. (2003). Intensity versus identity coding in an olfactory system. Neuron *39*, 991-1004.

Supekar, K., Menon, V., Rubin, D., Musen, M., and Greicius, M.D. (2008). Network analysis of intrinsic functional brain connectivity in Alzheimer's disease. PLoS computational biology *4*, e1000100.

Sutton, C., and McCallum, A. (2012). An introduction to conditional random fields. Foundations and Trends® in Machine Learning *4*, 267-373.

Tang, K., Ruozzi, N., Belanger, D., and Jebara, T. (2016). Bethe Learning of Graphical Models via MAP Decoding. Artificial Intelligence and Statistics (AISTATS).

van den Heuvel, M.P., Stam, C.J., Boersma, M., and Hulshoff Pol, H.E. (2008). Small-world and scale-free organization of voxel-based resting-state functional connectivity in the human brain. NeuroImage *43*, 528-539.

Wang, J., Zuo, X., and He, Y. (2010). Graph-based network analysis of resting-state functional MRI. Frontiers in systems neuroscience *4*, 16.

Yatsenko, D., Josic, K., Ecker, A.S., Froudarakis, E., Cotton, R.J., and Tolias, A.S. (2015). Improved estimation and interpretation of correlations in neural circuits. PLoS computational biology *11*, e1004083.

Yu, S., Huang, D., Singer, W., and Nikolic, D. (2008). A small world of neuronal synchrony. Cerebral cortex (New York, NY : 1991) *18*, 2891-2901.

Zuo, X.N., Ehmke, R., Mennes, M., Imperati, D., Castellanos, F.X., Sporns, O., and Milham, M.P. (2012). Network centrality in the human functional connectome. Cerebral cortex (New York, NY : 1991) *22*, 1862-1875.

**STAR methods**

**KEY RESOURCES TABLE**

Code used for CRF models can be found at https://github.com/hanshuting/graph\_ensemble.

**CONTACT FOR REAGENT AND RESOURCE SHARING**

Further information and requests should be directed to and will be fulfilled by corresponding authors.

**EXPERIMENTAL MODELS AND SUBJECT DETAILS**

**Animals and surgery**

All experimental procedures were carried out in accordance with the US National Institutes of Health and Columbia University Institutional Animal Care and Use Committee and have been described previously ([Carrillo-Reid et al., 2016](#_ENREF_13)). Briefly, simultaneous two-photon imaging and two-photon optogenetic experiments were performed on C57BL/6 male mice. Virus AAV1-syn-GCaMP6s-WPRE-SV40 and AVVdj-CaMKIIa-C1V1(E162T)-TS-P2A-mCherry-WPRE were injected simultaneously into layer 2/3 of left primary visual cortex (2.5 mm lateral and 0.3 mm anterior from the lambda, 200 μm from pia). After 3 weeks mice were anesthetized with isoflurane (1-2%) and a titanium head plate was attached to the skull using dental cement. Dexamethasone sodium phosphate (2 mg/kg) and enrofloxacin (4.47 mg/kg) were administered subcutaneously. Carprofen (5 mg/kg) was administered intraperitoneally. After surgery animals received carprofen injections for 2 days as post-operative pain medication. A reinforced thinned skull window for chronic imaging (2 mm in diameter) was made above the injection site using a dental drill. A 3-mm circular glass coverslip was placed and sealed using a cyanoacrylate adhesive ([Drew et al., 2010](#_ENREF_18)). Imaging experiments were performed 7~28 days after head plate fixation. During recording sessions mouse is awake (head fixed) and can move freely on a circular treadmill.

**Visual Stimulation**

Visual stimuli were generated using MATLAB Psychophysics Toolbox and displayed on a LCD monitor positioned 15 cm from the right eye at 45° to the long axis of the animal. Population activity corresponding to two-photon stimulation of targeted neurons in layer 2/3 of visual cortex was recorded with the monitor displaying a gray screen with mean luminescence similar to drifting-gratings. The imaging setup and the objective were completely enclosed with blackout fabric and a black electrical tape. Visual stimuli consisted of full-field sine wave drifting-gratings (100% contrast, 0.035 cycles/°, 2 cycles/sec) drifting in two orthogonal directions presented for 4 sec, followed by 6 sec of mean luminescence.

**Simultaneous two-photon calcium imaging and photostimulation**

Two-photon imaging and optogenetic photostimulation were performed with two different femtosecond-pulsed lasers attached to a commercial microscope. An imaging laser (λ = 940 nm) was used to excite a genetically encoded calcium indicator (GCaMP6s) while a photostimulation laser (λ = 1064 nm) was used to excite a red shifted opsin (C1V1) that preferentially responds to longer wavelengths ([Packer et al., 2012](#_ENREF_41)).

The two laser beams on the sample are individually controlled by two independent sets of galvanometric scanning mirrors. The imaged field of view was ~240X240 μm (25X NA 1.05 XLPlan N objective), comprising 60-100 neurons.

We adjusted the power and duration of photostimulation such that the amplitude of calcium transients evoked by C1V1 activation mimic the amplitude of calcium transients evoked by visual stimulation with drifting-gratings. Single cell photostimulation was performed with a spiral pattern delivered from the center of the cell to the boundaries of the soma at 0.001 pix/s for one second ([Carrillo-Reid et al., 2016](#_ENREF_13)).

**Image processing**

Image processing was performed with Image J (v.1.42q, National Institutes of Health) and custom made programs written in MATLAB as previously described. Acquired images were processed to correct motion artifacts using TurboReg . Neuronal contours were automatically identified using independent component analysis and image segmentation ([Mukamel et al., 2009](#_ENREF_39)). Calcium transients were computed as changes in fluorescence: (Fi – Fo)/Fo, where Fi denotes the fluorescence intensity at any frame and Fo denotes the basal fluorescence of each neuron. Spikes were inferred from the gradient (first time derivative) of calcium signals. We constructed an *N* x *F* binary matrix, where *N* denotes the number of active neurons and *F* represents the total number of frames for each movie. Peaks of synchronous activity describe population vectors ([Carrillo-Reid et al., 2008](#_ENREF_12)).

**Population analysis**

The method to identify neuronal ensembles that represent groups of neurons responding to a given visual stimuli was previously published ([Carrillo-Reid et al., 2015b](#_ENREF_11)). Briefly, to identify neuronal ensembles we constructed multidimensional population vectors that represent the simultaneous activation of different neurons. We tested the significance of population vectors against the null hypothesis that the synchronous firing of neuronal pools is given by a random process ([Carrillo-Reid et al., 2015a](#_ENREF_10); [Shmiel et al., 2006](#_ENREF_51)). Such population vectors can be used to describe the network activity as a function of time ([Brown et al., 2005](#_ENREF_7); [Carrillo-Reid et al., 2008](#_ENREF_12); [Sasaki et al., 2007](#_ENREF_46); [Schreiber et al., 2003](#_ENREF_48); [Stopfer et al., 2003](#_ENREF_56)). The number of dimensions for each experiment is given by the total number of active cells.

The similarity index between a pair of vectors is defined by their normalized inner product ([Carrillo-Reid et al., 2008](#_ENREF_12); [Sasaki et al., 2007](#_ENREF_46); [Schreiber et al., 2003](#_ENREF_48)), which represents the cosine of the angle between two vectors. Neuronal ensembles are defined by the concomitant firing of neuronal groups at different times. To identify the time course of each neuronal ensemble we obtained the singular value decomposition (SVD) of the significant patterns matrix *S*. In our particular case, the matrix *S,* that contains the information of the angles between all possible pairs of population vectors, is a real symmetric matrix given by S = V Σ VT, where V and VT represent orthonormal basis, and the elements of Σ are the singular values. The strongest factors from the SVD factorization define the number of cortical ensembles and their activation pattern ([Carrillo-Reid et al., 2015b](#_ENREF_11)).

**METHODS DETAILS**

**Allen Institute Brain Observatory dataset**

Population calcium recordings and changes in fluorescence were obtained using the SDK provided by Allen Institute.

**Conditional Random Fields**

We constructed conditional random fields (CRFs) as previously published ([Tang et al., 2016](#_ENREF_59)), using indicator feature vectors where , for each edge and node, and target binary population activity vectors , where , for samples (time points). For each sample, the conditional probability can be expressed as:

where is a vector of sufficient statistics of the distribution, is a vector of parameters, and is the partition function:

The conditional probability can be factored over a graph structure , where is the collection of nodes representing observation variables and target variables, and is the collection of subsets of . The conditional dependencies can be then written as

This model is a generalized version of Ising models, which have been previously applied to model neuronal networks ([Yu et al., 2008](#_ENREF_63)). The log-likelihood of each observation can be then written as:

Given the inferred binary spikes from raw imaging data, we construct a CRF model by two steps: (1) structure learning, and (2) parameter learning. For structure learning, we learned a graph structure using -regularized neighborhood-based logistic regression ([Ravikumar et al., 2010](#_ENREF_44)):

,

where

.

Here is a regularization parameter that controls the density of constructed structure. Then, a graph structure is learnt by thresholding the edge potentials with a given density preference . Edges with potential values within top quantile were kept as the final structure. It is worth noting that although could bias the result, varying does not lead to density values that differ much. This is probably because of the sparse nature of the obtained Ising model.

Based on the learned structure, we use the Bethe approximation to approximate the partition function, and iterative Frank-Wolfe methods to perform parameter estimation by maximizing the log-likelihood of the observations with a quadratic regularizer ([Tang et al., 2016](#_ENREF_59)):

Here is a regularization that controls the learnt parameters. Cross-validation was done to find the best , and via model likelihood. We varied with 6 values between 0.002 and 0.5, d with 6 values between 0.25 and 0.3, and with 5 values between 10 and 10000, all sampled uniformly. To obtain the best model parameters, 90% data were used for training, while 10% data were withheld for cross-validation. The best model parameters were determined by calculating the likelihood of the withheld data and selecting the parameter set with a locally maximum likelihood in the parameter space.

**Node strength**

We define the node strength as the sum of the ‘11’ term of edge potentials from all connecting edges:

Here denotes the number of connecting edges for node . The defined node strength reflects the importance of a given cell in co-activating with other cells.

**Shuffling method**

To generate shuffled models, we first randomize the spike raster matrices while preserving the activity per cell and per frame. Then, we trained CRF models using the shuffled spike matrices, with the cross-validated , and from the real model. This procedure is repeated 100 times. Random level of node strength is determined by mean ± S.D. of mean node strengths from all shuffled models.

**Identifying the most representative cortical ensembles**

To find the most representative cortical ensembles for each condition, we iterate through all the neurons and identify their contribution in predicting stimulus conditions with the population. To this end, for the neuron in population, we set its activity to be ‘1’ and ‘0’ in turn, in all M frames. With the two resulting population vectors in the frame among all samples, we calculate the likelihood of them coming from the trained CRF model:

Then, we computed the likelihood difference vector

and calculated the standard receiver operating characteristic (ROC) curve with the ground truth as the timing of each presented visual stimuli. The prediction ability of all nodes for all presented stimuli is then represented by an area under curve (AUC) matrix , where represents the AUC value of node predicting stimulus . Additionally, we calculated the node strength of each neuron in the CRF model.

The core ensemble for stimulus is defined by the following procedure: (1) find the nodes that has maximum value at column ; (2) vary a threshold between 0.5 and 0.9 with a step size of 0.05, and take all the nodes in (1) that has AUC values larger than the threshold; (3) take the population vector of the resulting nodes, calculate the cosine similarity with the binary data, predict timing of stimulus from the cosine similarity, and calculate accuracy of prediction; (4) threshold AUC values with the cross-validated threshold from (3) that results in the best accuracy; (5) shuffle the spike matrix for 100 times while preserving the activity per neuron and per frame, and train separate CRF models on shuffled data; (6) threshold the node strength of all nodes from (4) by the mean plus standard deviation of node strength from shuffled models; (7) the final representative cells are the remaining cells.

**Graph properties**

Given the adjacency matrix where if node is linked to node , we investigated the following graph properties: graph density, node degree, local clustering coefficient, and eigenvector centrality.

Graph density is calculated as the number of existing edges divided by the number of total possible edges:

where *NV* is the number of vertices in the graph.

Node degree is defined for node as the number of edges connected to it:

.

Local clustering coefficient is defined for each node as the fraction edges connected to it over the total number of possible edges between the node's neighbors (nodes that have a direct connection with it).

Eigenvector centrality is defined on the relative centrality score matrix , where

This can be written in the form of eigenvector equation:

Solving the above equation gives a set of eigenvalues and associated eigenvectors. The entry of the eigenvector associated with the largest gives the eigenvector centrality for the node.

**QUANTIFICATION AND STATISTICAL ANALYSIS**

CRF models were trained using the Columbia Yeti shared HPC cluster. MATLAB R2016a (MathWorks) were used for data analysis. Statistical details of each specific experiment can be found in figure legends.

**DATA AND SOFTWARE AVAILABILITY**

To demonstrate the general applicability of our approach we analyzed a publicly available dataset from the Allen Brain Observatory (<http://observatory.brain-map.org/visualcoding>) along with the SDK for extracting related information (<http://alleninstitute.github.io/AllenSDK/>) by Allen Institute of Brain Science. The experiments IDs are: 511507650, 511509529, 511510650, 511510670, and 511510855.

Other data and software used are available upon request.

**Supplementary Information**

**Figure S1. Cortical ensembles as a representation of multidimensional population vectors**

(A) Schematic representation of population vectors. Each point in a multidimensional space represents a population vector defined by a coactive group of neurons (left). A neuronal ensemble is defined by a cluster of population vectors (right). Related to Figure 1.

**Figure S2. CRF models with added nodes recapitulate the same properties of baseline CRF models**

(A) CRF graphs of baseline model (no added nodes) and the added node model, trained with the same experimental data. Edge color represents strength of node potentials (); node color represents node strength. Node size represents node degree. Scale bar: 50μm. (B) Graph density (P=0.4375 n.s), (C) node strength (P=0.8438 n.s), (D) node degree (P=0.1563 n.s), (E) clustering coefficient (P=0.1563 n.s) and (F) centrality (P=0.5625 n.s) values between baseline and added node models (n = 6 mice; Wilcoxon signed rank test). Each dot in the data plots represents the mean value for a different mouse. Related to Figure 2.

**Figure S3. Core ensembles from CRF have the highest classification performance for visual stimuli**

(A) Accuracy, (B) precision and (C) recall for classification performance from randomly down-sampled or up-sampled CRF core ensembles. (D) Accuracy, (E) precision and (F) recall for classification performance from randomly chosen ensembles. Data presented as box and whisker plots displaying median and interquartile ranges (n = 6 mice; Wilcoxon rank sum test). Related to Figure 4.

**Figure S4. Classification performance between CRF core ensembles and high OSI cells**

(A) ROC curves of classification performance between CRF core ensembles (orange) and high OSI cells (grey). Dashed line represent random performance. (B) AUC (P=0.9461 n.s), (C) accuracy (P=0.0001\*\*\*), (D) precision (P=0.0002\*\*\*) and (E) recall (P=0.3104 n.s) of classification performance for CRF core ensembles (orange) and high OSI cells (grey). (F) Percentage of shared neurons between CRF core ensembles and high OSI cells. (G) Percentage from the total population size of CRF core ensemble neurons and high OSI cells (P=0.5075 n.s). (H) Spatial maps of CRF core neurons (orange) and high OSI cells (grey). Circles represent neurons of horizontal (left) and vertical (right) visual stimuli, respectively. Neurons shared between CRF and high OSI cells are represented by orange dots circled by grey. Scale bar: 50μm. Data presented as box and whisker plots displaying median and interquartile ranges (n = 6 mice, 20 ensembles; Wilcoxon rank sum test). Related to Figure 5.

**Figure S5. High-ranked neurons from CRF models have pattern completion capability**

(A) CRF graphical model from simultaneous two-photon imaging and two-photon optogenetic single cell stimulation after imprinting protocol. Squares depict added nodes representing stimulation trials from 6 different neurons. Edge color tone represents edge potential strength (); node color represents node strength. Scale bar: 50μm. (B) High-ranked neurons (red) with pattern completion capability defined by AUC values and node strength. AUC values were calculated by predicting the photostimulation time of targeted neurons with single cell resolution using the cosine similarity function of core ensemble population vector from imprinted ensemble. Neurons with pattern completion capability are depicted in red. Neurons without pattern completion capability are depicted in blue. Dashed line and grey region represent random groups of neurons. (C) AUC values (P=0.0357\*) and (D) node strength (P=0.0357\*) of neurons with (red) and without (blue) pattern completion capability inferred with CRF graphical models. Note that two-photon optogenetic photostimulation of high-ranked neurons (red) was able to recall imprinted ensembles whereas single cell stimulation of non-high-ranked neurons (blue) was not able to induce pattern completion of imprinted ensembles. Data presented as box and whisker plots displaying median and interquartile ranges (n=1 mouse; Wilcoxon rank sum test). Related to Figure 6.

**Figure S6. Global network properties of cortical microcircuits remain stable after two-photon optogenetic stimulation**

(A) Node strength (P=0.9733 n.s), (B) node degree (P=0.2921 n.s), (C) clustering coefficients (P=0.7022 n.s) and (D) centrality (P=0.4225 n.s) values of non-stimulated neurons from Figure 6. Note that the whole network remains stable after ensemble imprinting. Data plots represent photostimulated neurons activated by imprinting protocol. (n=1 mouse, 74 neurons; Wilcoxon signed rank test)