Identification and Targeting of Cortical Ensembles with Probabilistic Graphical Models

Luis Carrillo-Reid¹*, Shuting Han¹*, Ekaterina Taralova¹,², 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 author:
Luis Carrillo-Reid. Ph.D. lc2998@columbia.edu
906 NWC Building 550 West 120 Street, Box 4822, New York, NY, 10027.

* These authors contributed equally.

Summary

Because the neural code is likely a result of the emergent properties of neural circuits, a fundamental problem in neuroscience is the characterization of physiological and behavioral correlates of neuronal ensemble activity. Recent developments in optical techniques have allowed the recording and manipulation of neuronal populations with single-cell precision. However, to fully understand how the activity of specific neuronal populations allows an internal representation of the external world, it is
necessary to generate empirically grounded models that capture changes in ensemble
dynamics. We used Conditional Random Fields (CRFs), a structured prediction
method often applied in pattern recognition, to identify cortical ensembles and
describe the reconfiguration of network dynamics induced by population
photostimulation of neurons from mouse primary visual cortex in vivo. CRFs allowed
us not only to identify the most influential neurons in cortical ensembles, but also to
target individually core neurons that are functionally key for pattern completion.
These results demonstrate the potential of machine learning techniques to
characterize neuronal population responses and network dynamics in cortical
microcircuits.

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
1, 2, 3, 4, 5, 6, 7. Recent advances in two-photon calcium imaging and two-photon optogenetics
have made possible the simultaneous reading and writing of cortical ensemble activity with
single cell resolution in awake animals 8. However, how the activation of specific groups
of neurons relates to functional or behavioral changes has been difficult to elucidate,
because it requires online identification of core neurons that can be targeted during
optogenetic experiments.

One suitable method to identify core neuronal ensembles is graph theory. Indeed, graph
theory has been used in neuroscience to describe the structural and functional organization
of entire brains 9. However, such graphs are usually constructed with nodes representing
brain regions, and edges representing information flow. For functional analysis, many studies have constructed graphs with data from fMRI, EEG and electrode arrays, taking brain regions, voxels or electrode position as nodes, and activity associations such as cross correlation, mutual information and Granger causality as edges. In addition, at the single cell level, graphs have been used to describe organizing principles of artificial neural networks. Such graphs are usually associated with a restricted set of parameters that describe the weight and direction of edges obtained by pairwise metrics, therefore have limitations for characterizing changes in overall network dynamics and properties underlying population activity. Finally, a few studies have applied graph theory to describe network organization in calcium imaging data with single cell resolution in cultures or brain slices. Nevertheless, these methods have been applied only as a descriptive tool. To define neuronal ensembles and their dynamic reconfiguration due to external stimuli in awake animals it is necessary to identify and manipulate with high spatial precision neurons that could have a potential role orchestrating the overall network activity.

Cortical ensembles in primary visual cortex consist of coactive neurons, forming a network structure that can be intuitively characterized with probabilistic graphical models, where nodes and edges are biologically meaningful, representing neurons and their functional connections respectively. Here, we exploit probabilistic graphical models to analyze two-photon calcium imaging recordings of mouse primary visual cortex. More specifically, we consider CRFs, a combination of graph theory and probabilistic modeling. CRFs use graphs to simplify and express the conditional independence structure between a collection of random variables. We demonstrate that CRF models allow the
identification of cortical ensembles associated with different experimental and physiological conditions, and the prediction of the neurons that are most efficient at pattern completion. This method opens the possibility of targeting, with single cell precision, the core neurons from specific populations, to specifically alter microcircuit function. Importantly, as opposed to dimensionality reduction methods, CRFs generate a functional model of the circuit, one that can be used to probe and explore a host of biological questions.

Results

CRF models can encode population responses of visual cortex to oriented stimuli

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 $^{28}$. 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 $^{29}$, bioinformatics $^{30,31,32}$, computer vision $^{33,34}$ and natural language processing $^{35,36}$.

In order to study the network dynamics 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 (Fig. 1a). We used drifting-gratings as visual stimuli. Population vectors representing the coordinated activity of neuronal groups were inferred from
calcium imaging recordings \(^{37}\) and used as training data (Fig. 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 through connecting edges, which encode four possible configurations ‘00’, ‘01’, ‘10’, and ‘11’, depending on the values of the two nodes the edge is adjacent to. The two values associated with nodes and the four values associated with edges are characterized by a set of parameters called node potentials \((\phi_0, \phi_1)\) and edge potentials \((\phi_{00}, \phi_{01}, \phi_{10}, \phi_{11})\) respectively (Fig. 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 (e.g. training data), we estimated the model’s structure (e.g. graph connectivity) and parameters (e.g. the potentials) via maximum likelihood. We then performed cross-validation on held-out data to identify the optimal level of sparsity and regularization (see Experimental Procedures). Once the model is learned from the data, the normalized product of the corresponding nodes and edge potentials describes the probability distribution of a neuronal population that represents a specific activation pattern. This is a normalized probability over all possible network states (e.g. all binary configurations of the nodes) and is estimated by maximizing likelihood on past observed population vectors (Methods).

To integrate information about 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 (Fig. 1c-d). The general and mathematical properties of CRF models obtained with added nodes did not significantly differ from CRF models obtained without added nodes.
(Supplementary Fig. 1). 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 model visual stimuli from observed data. In this way, the nodes directly connected to the added nodes represent a specific model for different visual stimuli (Fig. 1d). Given two different visual stimuli (horizontal or vertical drifting gratings), the logarithm of probability corresponding to observing each stimulus is defined by \( p_h = p(x^m|y^m, y_h = 1, y_v = 0) \) and \( p_v = p(x^m|y^m, y_h = 0, y_v = 1) \). Thus, assuming that the a priori probability of visual stimuli is uniform, the higher probability value through comparing the log likelihood ratio \( \text{argmax} \log(p_h) - \log(p_v) \) can be used to classify the presented stimuli (Fig. 1e).

These results demonstrated the ability of CRFs to accurately model different orientations of drifting-gratings from calcium imaging population recordings.

Identification of core neurons in cortical ensembles with CRFs

Cortical ensembles represent coactive neuronal populations that could represent the computational unit of brain functions. The structural and functional organization of brain microcircuits could be characterized by modular properties, where a group of core neurons might orchestrate the overall network dynamics. For this reason, in order to design close-loop optogenetic experiments with single cell resolution, it is necessary to identify these cortical modules formed by “core neurons” that can efficiently represent different physiological functions from cortical ensembles. To identify core neurons from cortical 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 logarithm probability difference (probability ratio) \( \log p_{t,1}^m - \log p_{t,0}^m = \log p(y^m|x_t^m, x_i^m = 1) - \log p(y^m|x_t^m, x_i^m = 0) \) using the inferred CRF models (Fig. 2a). Then, we calculated single neuron performance by binarizing the logarithm probability difference (Fig. 2b) and calculating the area under the curve (AUC) from the receiver operating characteristic curve (ROC). Since core neurons are likely to have concomitant activity, we used the node strength obtained from CRFs models that represents the summation of edge potentials (\( \phi_{11} \) terms) from all connecting edges for each node in the graph. In this way, highly connected neurons or strongly connected neurons have high node strength whereas weakly connected neurons have low node strength (Fig. 2c). Finally, 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 (Fig. 2d).

To demonstrate the general applicability of our method, we analyzed publically open datasets (Allen Brain Observatory Release 2016) 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 was able to find core ensembles for each visual stimulus (Fig. 3; overall classification AUC for TF = 1: 0.9129 ± 0.0134 [S.D.]). Core ensembles allowed us to demonstrate that classification performance to different drifting-gratings in layer 2/3 of primary visual cortex was significantly better for lower temporal frequencies, as measured by ROC curves and AUC values (Fig. 3c-e; classification AUC: \( AUC_{TF=1} = 0.9129 \pm 0.0134, AUC_{TF=2} = 0.8973 \pm 0.0141, AUC_{TF=4} = 0.8063 \pm 0.0226, AUC_{TF=8} = 0.7045 \pm 0.0262, AUC_{TF=15} = 0.6517 \pm 0.0171; p_{1,4}<0.01, p_{1,8}, p_{1,15}<0.001; n = 5 \) animals, 20 ensembles; Wilcoxon rank sum test).
Prediction performance of core neurons

We next investigated whether these core neurons from cortical ensembles were optimal for predicting visual stimuli. To do so, we randomly shuffled 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 had a maximum value when the population vectors were unchanged (Fig. 4a-c, orange; original core neurons similarity 0.2887 ± 0.0926 [S.D.]; prediction AUC 0.9383 ± 0.0333).

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 metrics, population vectors from core ensembles achieve the best accuracy, precision and recall when predicting the presented visual stimuli, compared with resized ensembles (Supplementary Fig. 2a-c, orange; original core neurons prediction accuracy 0.8367 ± 0.0623, precision 0.6175 ± 0.1752, recall 0.9067 ± 0.0717).

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 raises the question of whether such population vectors are a specific non-random subgroup. To answer this, we randomly sampled a subset from all the population of neurons ranging from 10% to 190% of the total size of core ensembles. Indeed, we observed that the prediction performance from random groups of neurons was significantly
lower than the core ensembles identified by the CRF models (Figures 3a-c, black; Supplementary Fig. 2a-c, black; best similarity 0.1985 ± 0.0590 [S.D.], AUC 0.5029 ± 0.0616, accuracy 0.6906 ± 0.0671, precision 0.2570 ± 0.1756, recall 0.3617 ± 0.1027), indicating that the core neurons from cortical ensembles achieved a classification performance considerably better than random sets of neurons.

Identification of pattern completion neurons using CRFs models

One major advantage of CRFs is their ability to construct graphical models that capture network properties that could be used to target individual neurons. We therefore decided to exploit this advantage by targeting neurons with optogenetics in order to alter circuit function. It has been recently shown that the repetitive activation of an identified neuronal population with two-photon optogenetics imprints an artificial cortical ensemble that can be recalled later on by activating specific members of the ensemble \(^8\). Since CRFs can be used to identify core neurons from cortical ensembles evoked by visual stimuli, we perform a similar approach for the identification of neurons with pattern completion capability from artificially imprinted cortical ensembles. To do so, we used structural and classification parameters of CRFs learned from simultaneous two-photon imaging and two-photon optogenetic experiments with single cell resolution (Supplementary Fig. 3a) to define high-ranked neurons as the ones with strong node strength and high AUC values (Supplementary Fig. 3b). Our experiments demonstrated that single-cell two-photon optogenetic stimulation of high-ranked neurons (AUC 0.8397 ± 0.0361; node strength -0.1405 ± 0.0770) 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) were unable to recall imprinted cortical ensembles (Supplementary Fig. 3c-d). These experiments support the hypothesis that neurons with pattern completion capability represent neurons that are highly connected with other members of artificially imprinted cortical ensembles.

CRFs models reveal the reconfiguration of cortical microcircuits

Another advantage of CRF models could be the ability to capture changes in network connectivity under different experimental conditions. To test this we compared CRF models estimated from data before and after two-photon population manipulation of a given set of neurons for several times (Fig. 5a), an experimental protocol that reconfigures network activity building new coactive ensembles. To visualize the change in graph connectivity induced by the imprinting protocol in neurons with pattern completion capability we constructed isomorphic graphs from the CRF models and arranged them in a circular visualization (Fig. 5b). After the artificial ensemble was imprinted, neurons with pattern completion capability showed better predictive performance and higher node strength (Fig. 5c; stimulated neuron: node strength increased from -0.1539 to 0.1609, prediction AUC increased from 0.5234 to 0.7447). This demonstrated that structural and classification parameters inferred with CRFs can be used to study changes in network properties of specific neurons and those very CRFs can also be used to target single neurons that play a key role in the computational properties of cortical microcircuits. Interestingly, for non-photostimulated neurons, the graph properties of CRFs before and after population...
photostimulation remained stable (Supplementary Fig. 4) suggesting that imprinted
ensembles have been added to cortical microcircuits while preserving a balance with the
overall network structure.
Machine learning analysis of functional connectivity in cortical microcircuits

In this study, we provide a novel tool for the identification and targeting of key members of cortical ensembles from population calcium imaging recordings of mouse primary visual cortex \textit{in vivo} using Conditional Random Fields (CRFs), a probabilistic and graphical machine learning method. As opposed to traditional descriptive approaches for neuronal ensemble identification and network analyses based on correlations between pairs of neurons, machine learning methods represent an empirically grounded approach to create models that aim to capture the functional structure of neuronal circuits and also provide information about the network properties of individual neurons within a population.

In the past decades, graph theory has been applied to characterize the structure and function of neuronal networks\textsuperscript{11, 12, 13, 14, 17, 18, 40, 41, 42, 43, 44}. While most of these studies operated on functional recordings across multiple brain regions\textsuperscript{12, 13, 17, 40, 45}, only a few have focused on the general network properties of cortical circuits with recordings from single neurons\textsuperscript{23, 25, 39, 46}. The majority of methods applied to infer network properties in brain slices\textsuperscript{1, 2, 39, 46, 47} or \textit{in vivo}\textsuperscript{25} operate on the correlation matrix, and aim to recover the functional dependencies between observed neurons. 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. The generation of a graphical model of the circuit has obvious advantages, since it provides a direct link to its functional...
connectivity structure, enabling targeted investigations and manipulations of its biological components.

CRFs graphical models identify neuronal ensembles

Compared with fully generative models such as Markov Random Fields and Bayesian Networks 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. 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. Therefore, CRFs appear to be promising for modeling cortical functional connectivity and for identifying core ensembles that could be easily manipulated by two-photon optogenetics.

The computational difficulty in constructing CRFs lies in recovering the global normalizer (the partition function) and gradients of global normalizer. With an arbitrary graph structure, this problem is often intractable. But 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 \(^{48}\). Thus, CRFs can in principle be applied to datasets with thousands of interconnected neurons. However, for datasets with more neurons (and therefore more random variables and larger networks), CRFs (like most machine learning approaches) would require an increasingly large 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 location 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) \(^{37,38}\). Even though SVD can identify cortical ensembles reliably, it lacks a structured graphical model that allows the systematic study of changes in network properties. One limitation for the current CRF learning algorithms is the computation of a sparse graph structure which is less prone to over-fitting of the data. Certainly, due to some modeling assumptions, computational considerations and the finiteness of training data, the learned graphical structure and parameters may not be the globally best probabilistic model of the data. However, it is recovered efficiently rather than requiring an exhaustive and unrealistic exploration of 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

The key advantage of CRF models is their ability to model the circuit explicitly, in a manner than can be used for targeted manipulation. In visual cortex, non-targeted electrical stimulation has been used for decades as an attempt to provide useful visual sensations to patients that have lost the functionality of their eyes \(^49\). The sensations produced by electrical stimulation of the visual cortex were termed phosphenes since they represented bright spots. To improve prostheses, one could in principle train patients using devices with a large number of electrodes \(^50\). Our results suggest that after a given network have been imprinted \(^8\), the identification of neurons with pattern completion capability could be used to reduce the number of active points that require stimulation and pinpoint them with surgical accuracy. 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.

Finally, it has been shown that the connectivity of diverse systems described by graphs with complex topologies follow a scale-free power-law distribution \(^51\). Scale-free networks are characterized by the existence of a small subset of nodes with high connectivity \(^38\). Similarly, core 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 \(^8, 14, 23\). However, the efficient identification of such neurons from network models that capture the overall properties of neuronal microcircuits was lacking.
In closing, our results suggest that the structural and predictive parameters defined by CRFs 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.

Methods

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. 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 postoperative 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. 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. 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.

**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. Calcium transients were computed as changes in fluorescence: \( \frac{F_i - F_o}{F_o} \), where \( F_i \) denotes the fluorescence intensity at any frame and \( F_o \) denotes the basal fluorescence of each neuron. Spikes were inferred from the gradient (first time derivative) of calcium signals with a threshold of 3 S.D. above noise level. We constructed an \( N \times F \) 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.

**Population vectors**

We constructed multidimensional population vectors that represent the simultaneous activation of different neurons., only high-activity frames were used. An activity threshold was determined by generating 1000 shuffled raster plots and comparing the distribution of the random peaks against the peaks of synchrony observed in the real data. We tested
the significance of population vectors against the null hypothesis that the synchronous
firing of neuronal pools is given by a random process \(^{37,38,56}\). Such population vectors can
be used to describe the network activity as a function of time \(^{55,57,58,59,60}\). 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 then defined by their normalized inner product \(^{55,58,60}\),
which represents the cosine of the angle between two vectors. Neuronal ensembles are
defined by the concomitant firing of neuronal groups at different times.

**Allen Institute Brain Observatory dataset**

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 fluorescence and dF/F (http://alleninstitute.github.io/AllenSDK/) by Allen Institute of Brain Science. Spikes were
detected by first low-pass filtering the dF/F traces, then a threshold of 5 S.D. above noise
level on the first derivative of filtered dF/F. The experiments IDs used are: 511507650,
511509529, 511510650, 511510670, and 511510855.

**Conditional Random Fields**

We constructed conditional random fields (CRFs) as previously published \(^{48}\), using
indicator feature vectors \(\mathbf{x} = [x^1, x^2, \ldots, x^M]\), where \(x^m \in \mathcal{X}\), for each edge and node, and
target binary population activity vectors \(\mathbf{y} = [y^1, y^2, \ldots, y^M]\), where \(y^m \in \mathcal{Y}\), for \(M\)
samples (time points). For each sample, the conditional probability can be expressed as:
$$p(y^m|x^m; \theta) = \frac{\exp((\phi(x^m, y^m), \theta))}{Z(x^m; \theta)}$$

where $\phi$ is a vector of sufficient statistics of the distribution expressed in log-linear form, $\theta$ is a vector of parameters, and $Z$ is the partition function:

$$Z(x^m; \theta) = \sum_{y \in \mathcal{Y}} \exp((\phi(x^m, y), \theta))$$

The conditional probability can be factored over a graph structure $\mathcal{G} = (\mathcal{V}, \mathcal{A})$, where $\mathcal{V}$ is the collection of nodes representing observation variables and target variables, and $\mathcal{A}$ is the collection of subsets of $\mathcal{V}$. The conditional dependencies can be then written as

$$p(Y|X; \theta) = \frac{\exp(\sum_{i \in \mathcal{V}} \theta_i \phi_i(X, Y_i) + \sum_{a \in \mathcal{A}} \theta_a \phi_a(X, Y_a)}{Z(X; \theta)}$$

This model is a generalized version of Ising models, which have been previously applied to model neuronal networks\(^{41}\). The log-likelihood of each observation can be then written as:

$$\ell(\theta; X^m, Y^m) = (\phi(X^m, Y^m), \theta) - \log Z(X^m).$$

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 $\ell_1$-regularized neighborhood-based logistic regression\(^{61}\):

$$\min_{\theta \in \mathcal{V}} \{\ell(\theta; x) + \lambda \|\theta\|_1\},$$

where
\[ \ell(\theta; x) = -\frac{1}{n} \sum_{t=1}^{n} \log \frac{\exp(2x_r \sum_{t \in V \setminus r} \theta_{rt} x_t)}{\exp(2x_r \sum_{t \in V \setminus r} \theta_{rt} x_t) + 1} \]

\[ \theta_{\setminus r} = \{ \theta_{ru}, u \in V \setminus r \}. \]

Here \( \lambda_s \) is a regularization parameter that controls the sparsity (or conversely, the density) of the constructed graph structure. The final graph structure is obtained by thresholding the edge potentials with a given density preference \( d \). Edges with potential values within the top \( d \) quantile are kept as the final structure. It is worth noting that although \( d \) could bias the result, varying \( d \) does not lead to density values that differ much. This is because of the sparsity induced by the \( \ell_1 \) regularizer.

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 (equation 2) with a quadratic regularizer:

\[ \ell(\theta; X, Y) = \sum_{m=1}^{M} \ell(\theta; X^m, Y^m) - \frac{\lambda_p}{2} \| \theta \|^2 \]

Here \( \lambda_p \) is a regularization that controls the learnt parameters and helps preventing over-fitting. Cross-validation was done to find the best \( \lambda_s \), \( d \) and \( \lambda_p \) via held out model likelihood. We varied \( \lambda_s \) with 6 values between 0.002 and 0.5, \( d \) with 6 values between 0.25 and 0.3, and \( \lambda_p \) 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 withhold 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:

\[ s(i) = \sum_{j=1}^{N_E(i)} \phi_{11}(i,j). \]

Here \( N_E(i) \) denotes the number of connecting edges for node \( i \). 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 \( \lambda_s, d \) and \( \lambda_p \) 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 the stimulus conditions in the
population. To this end, for the \( i^{th} \) 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 $m^{th}$ frame among all samples, we calculate the log-probability of them coming from the trained CRF model using equation (1):

$$p_{i,1}^m = p(y^m|x_{\setminus i}^m, x_i^m = 1; \theta),$$

$$p_{i,0}^m = p(y^m|x_{\setminus i}^m, x_i^m = 0; \theta).$$

Then, we computed the log likelihood ratio:

$$l_{i,1-0} = \{\log(p_{i,1}^m) - \log(p_{i,0}^m)\}, m = 1, \ldots, M$$

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 $A$, where $A_{i,s}$ represents the AUC value of node $i$ predicting stimulus $d$. Additionally, we calculated the node strength $S = \{s_i\}$ of each neuron in the CRF model.

We then computed prediction AUC $A^r$ and node strength $S^r$ of each node from 100 CRF models trained on shuffled data. The final core ensemble for stimulus $d$ is defined as:

$$\{i|A_{i,s} > \text{mean}(A_s^r) + \text{std}(A_s^r), S_i > \text{mean}(S^r) + \text{std}(S^r)\}.$$

**Graph properties**

Given the adjacency matrix $A = (a_{v,t})$ where $a_{v,t} = 1$ if node $v$ is linked to node $t$, 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:

\[ d = \frac{\sum_{vt} a_{vt}}{N_V(N_V - 1)/2} \]

where \( N_V \) is the number of vertices in the graph.

Node degree is defined for node \( v \) as the number of edges connected to it:

\[ \text{deg}(v) = \sum_t a_{vt} \]

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 \( X = (x_v) \), where

\[ x_v = \frac{1}{\lambda} \sum_{t \in G} a_{vt} x_t \]

This can be written in the form of eigenvector equation:

\[ Ax = \lambda x \]

Solving the above equation gives a set of eigenvalues \( \lambda \) and associated eigenvectors. The \( v^{th} \) entry of the eigenvector associated with the largest \( \lambda \) gives the eigenvector centrality for the \( v^{th} \) node.
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. All numbers in the text and figure legends denote mean ± S.E.M. unless otherwise indicated.

Resource Availability

Code used in this paper can be found at https://github.com/hanshuting/graph_ensemble.

References


**Acknowledgments:** We thank 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.


Competing Financial Interests

The authors declare no competing financial interests.

Materials & Correspondence

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

906 NWC Building 550 West 120 Street, Box 4822, New York, NY, 10027.
Figure 1. Classification of visual stimuli from calcium imaging data 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) Schematic representation of population vectors. Each point in a multidimensional space represents a population vector defined by a coactive group of neurons. A neuronal ensemble is defined by a cluster of population vectors. (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) Graphical representation of connectivity with the added nodes. Added nodes (squares) represent two stimuli in this case: horizontal (red) and vertical (blue) drifting-gratings. Nodes that are connected with them correspondingly are highlighted in the same color. (e) Ratio of log 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 in 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 $i^{th}$ neuron is set to '1' or '0' at each frame, and the log likelihood $p_{i,1}^m$ and $p_{i,0}^m$ of modified population vectors is calculated. Edge color tone represents edge potential strength ($\phi_{11}$); node color represents node strength. (b) Log 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 predictive performance at low temporal frequencies (a) Temporal course of ensemble classification using CRFs. Colored stripes indicate different visual stimuli. Scale bar: 200 frames. (c) ROC curves of core ensembles from TF = 1 Hz predicting different temporal frequencies (TF: 1, 2, 4, 8 and 15 Hz) (d) AUC for different TFs compared to TF=1 Hz (classification AUC: $\text{AUC}_{\text{TF}=1} = 0.9129 \pm 0.0134$, $\text{AUC}_{\text{TF}=2} = 0.8973 \pm 0.0141$, $\text{AUC}_{\text{TF}=4} = 0.8063 \pm 0.0226$, $\text{AUC}_{\text{TF}=8} = 0.7045 \pm 0.0262$, $\text{AUC}_{\text{TF}=15} = 0.6517 \pm 0.0171$; $p_{1,2}=0.4249$ n.s; $p_{1,4}=0.0003^{***}$; $p_{1,8}=2.9598e-07^{***}$; $p_{1,15}=6.7956e-08^{***}$). 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 high 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 (orange), and randomly chosen neuron groups (black). Black triangle indicates the original core ensemble size. (b) AUC values from core population vectors from randomly down-sampled or up-sampled population vectors (orange), and randomly chosen neuron groups (black) predicting the time course of visual stimuli. Note that randomly removing or adding elements from the core ensemble decreases the ability to predict visual stimuli. (c) ROC curves of randomly down-sampled or up-sampled population vectors (orange), and randomly chosen neuron groups (black). Line color represents the size of ensembles as a ratio with the original core ensemble size. (n=6 mice, 20 ensembles; Wilcoxon rank sum test).
Figure 5. 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 ($\phi_{11}$); node color represents node strength. Node size represents node degree. Scale bar: 50\mu m.

(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. (n=15 neurons; Wilcoxon signed rank test).
Figure S1. 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 ($\phi_{11}$); node color represents node strength. Node size represents node degree. Scale bar: 50\(\mu\)m. (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 (g) 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.
Figure S2. 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).
Figure S3. 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 ($\phi_{11}$); 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 (Wilcoxon rank sum test).
Figure S4. 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=74 neurons; Wilcoxon signed rank test)