

FUNCTIONAL CONNECTIVITY IN CORTICAL CIRCUITS

A Dissertation Submitted to the Faculty of

The Graduate School

Baylor College of Medicine

In Partial Fulfillment of the

Requirements for the Degree

of

Doctor of Philosophy

by

Dimitri Yatsenko

Houston, Texas

December 2014

APPROVED BY THE DISSERTATION COMMITTEE

Signed _____

Andreas S. Tolias, Ph.D.

Chairperson

Signed _____

Fabrizio Gabbiani, Ph.D.

Signed _____

Krešimir Josić, Ph.D.

Signed _____

Daoyun Ji, Ph.D.

Signed _____

Athanassios G. Siapas, Ph.D.

APPROVED BY THE DEPARTMENT OF NEUROSCIENCE

Signed _____

Paul J. Pfaffinger, Ph.D.

Director of Graduate Studies

Signed _____

Dora E. Angelaki, Ph.D.

Chairman of Department

APPROVED BY THE DEAN OF GRADUATE SCIENCES

Signed _____

Hiram F. Gilbert, Ph.D.

Date _____

Acknowledgments

My six years of graduate studies at Baylor College of Medicine have been a great privilege, challenge, and adventure. During this time, I received tremendous support from my mentors, instructors, sponsors, collaborators, labmates, administrative staff, and friends.

I am profoundly grateful to my supervisor, Dr. Andreas S. Tolias for allowing me to become part of the multidisciplinary research environment that he has built in his lab. From our first interview and throughout the many projects we have undertaken since, he has encouraged me to think boldly yet critically. His optimism and enthusiasm made it possible to work through several uncertain phases of my projects and to define new goals. I thank him for the freedom to explore and pursue my own ideas.

I thank the past and current members of my dissertation committee Drs. Fabrizio Gabbiani, Daoyun Ji, Peter Saggau, Krešimir Josić, and Athanassios G. Siapas for their continued advice and guidance. In particular, Dr. Josić became closely involved in the definition and verification of the major mathematical aspects of this dissertation.

This work was performed in close collaboration with my lab mates. In particular, Emmanouil Froudarakis sharing his two-photon imaging data to be included in this dissertation. Cathryn R. Cadwell provided transgenic mice with fluorescently labeled interneurons that were used in this study. I also thank my former labmate Dr. R. James Cotton who, in collaboration with Prof. Peter Saggau, developed the unique two-photon

microscope for high-speed *in vivo* calcium imaging that enabled the level of analysis developed in this dissertation. I thank Alexander S. Ecker for critically examining the assumptions underlying my statistical models and helping eliminate several logical flaws early in the process.

The early stages of this work were supported in part by a Predoctoral Fellowship from the Theoretical and Computational Neuroscience Training Program of the Gulf Coast Consortia. The financial support and the excellent coursework and networking opportunities organized through this fellowship were instrumental in bringing this work to fruition. I am grateful to Dr. Valery Kalatsky, then at the University of Houston, who served as my co-mentor during this fellowship.

I also wish to express my gratitude to the BCM Neuroscience Department for running an excellent training program with great resources and opportunities available to its students.

Abstract

Neurons in sensory cortical areas are embedded in local microcircuits with their own dynamics and intrinsic patterns of activity. Yet present knowledge of the functional architecture of sensory circuits is formulated, primarily, in terms of the response properties of individual cells to external stimuli. Further understanding of cortical computation depends on characterizing the simultaneous activity of entire local circuits and analyzing interactions between their components, *i.e.* their *functional connectivity*. Therefore, a foremost challenge in systems neuroscience is to record the simultaneous spiking activity of sufficiently large and densely sampled populations, to extract the essential properties of such activity, and to relate these properties to circuit anatomy.

To address this challenge, we used high-speed 3D two-photon imaging of calcium signals in mouse primary visual cortex to record the simultaneous activity of nearly every neuron in volumes 200 μm wide and 100 μm deep (150–450 cells) during visual stimulation. Previous studies have characterized population calcium activity by using Pearson correlation matrices, cross-correlations, or coactivation patterns. We introduced the use of *pairwise partial noise correlations*, *i.e.* correlations between pairs of neurons remaining after subtracting the portions of their activity predicted linearly from the activity of all the other recorded cells as well as the external stimulus, for describing the functional connectivity in neuronal population signals. We hypothesized that such conditioning would reveal more immediate interactions between individual

cells and exhibit stronger relationships to circuit anatomy than Pearson correlations. First, we showed that estimation of partial noise correlations required *regularization*, *i.e.* imposing a structure on the estimate in order to reduce the estimation error. Among several regularization schemes that we evaluated, the optimal scheme decomposed the matrix of partial pairwise correlations into a sparse component and a low-rank component. Conceptually, the sparse component represents ‘interactions’ between specific pairs of the recorded neurons whereas the low-rank component represents ‘latent units’, *i.e.* widespread fluctuations such as arising from shared divergent input or from emergent synchronous activity.

To verify that partial correlations were effective expressions of functional connectivity, we demonstrated that they exhibited stronger relationships to circuit anatomy and to the functional properties of individual cells than Pearson correlations. For example, we found that they more clearly differentiated similarly tuned pairs from differently tuned pairs and nearby pairs from distant pairs. Furthermore, genetic labeling of parvalbumin-positive (PV+) interneurons revealed distinct levels of partial correlations between PV-/PV-, PV+/PV-, and PV+/PV+ pairs. Surprisingly, PV+/PV+ pairs had the strongest partial correlations with dense positive pairwise connectivity. This finding contradicts the commonly held belief that PV+ interneuronal networks reflect the common fluctuations of the local microcircuit. Instead, this finding is consistent with detailed, specific input from a spontaneously active layer such as L5. Furthermore, partial correlations of distantly spaced PV+/PV- pairs exhibited a consistently negative average partial correlations consistent with lateral inhibition of pyramidal cells by PV+ interneurons.

These results show that partial correlations in densely sampled groups of cortical neurons, regularized as a mixture of a sparse and a low-rank components, provide useful description of population activity during sensory processing as demonstrated by their

improved differentiation of anatomical attributes of the circuit. Furthermore, the relationships between the structure of the estimate and circuit anatomy suggested hypotheses for mechanistic interactions in the circuit, which can be tested by combining these studied with other anatomical techniques in future studies.

Contents

Acknowledgments	3
Abstract	5
1 Introduction	13
1.1 Computations in cortical microcircuits	14
1.2 Population calcium signals	15
1.3 Measures of functional connectivity	16
1.4 Main contribution: Functional connectivity through partial correlations .	17
1.5 Overview	18
2 Improved estimation and interpretation of correlations in neural circuits	19
2.1 Introduction	21
2.1.1 Simulation	31
2.2 Discussion	47
2.3 Materials and Methods	55
3 Differential functional connectivity of parvalbumin-positive interneurons	63
3.1 Introduction	64
3.2 Results	65
3.2.1 Networks of partial correlations	65

3.2.2	Functional connectivities by cell pair type	66
3.2.3	Distance dependence of functional connectivity by cell pair type . .	67
3.2.4	Orientation tuning dependence of functional connectivity by cell type	73
3.3	Discussion	73
3.4	Methods	75
3.4.1	Data acquisition and processing	75
3.4.2	Animals	76
4	Conclusions and future directions	77
	Bibliography	95

List of Figures

1	Regularized correlation matrix estimators	28
2	Performance of covariance estimators on samples from Ising models	34
3	Estimation of noise correlations from population calcium signals	36
4	Evaluation of correlation estimators by cross-validation	39
5	Optimization of regularization parameters	40
6	Example of sparse partial correlation structure	42
7	Properties of $C_{\text{sparse+latent}}$ estimates	44
8	Relationship between functional connectivity and circuit architecture	46
9	Examples of functional connectivity in populations with labeled PV+ interneurons	68
10	Functional connectivities for PV−/−, PV−/+ , and PV+/+ cell pairs	70
11	Distance dependence of functional connectivity by cell pair type	72
12	Orientation tuning dependence of functional connectivity by cell pair type	74

Chapter 1

Introduction

1.1 Computations in cortical microcircuits

Despite the vast amount of exquisite detail known about the anatomy and physiology of the cerebral cortex, its general operating principles remain unknown. Understanding these principles holds promise of treating neurological disorders and even replicating in artificial machines the kinds of cognitive abilities that only animals have so far possessed.

The central question addressed in this dissertation is the organization of interactions in large local groups of cortical neurons inferred directly from observations of their activity. With technologies allowing to record the simultaneous activity of large three-dimensional clusters of cells (up to 400 cells) *in vivo* with high temporal resolution, we examined patterns of interactions with respect to physical distances separating the cells as well as the cell types and the stimulus response properties of individual cells. These characteristics of interactions are key factors in constraining the search for the long-hypothesized canonical organization of cortical computations.

Several features of cortical architecture suggest that it may have a modular, uniform functional design — an array of smaller modules carrying out canonical, general-purpose computations versatile enough to solve different problems in different brain regions (Mountcastle, 1997; Douglas and Martin, 2007). Within the cortex, local synaptic connectivity is dense and recurrent while different cortical areas exhibit similar connectivity patterns between cortical layers, between different cell types, and to and from the subcortical regions (Douglas and Martin, 2004; Harris and Mrsic-Flogel, 2013). The multineuronal spiking activity of local populations consists of stereotyped dynamical patterns that are largely shaped by the local recurrent networks and can be triggered by external stimuli or occur spontaneously (MacLean et al., 2005; Luczak et al., 2009; Harris et al., 2011; Hofer et al., 2011; Miller et al., 2014). Additionally, cortical microcircuits can

support multiple information processing streams with distinct computations and projection targets (Vélez-Fort et al., 2014). If local canonical computations do indeed exist, describing and understanding the repertoire of their functions could hold the key to a broad understanding of brain function.

1.2 Population calcium signals

Functional characterization of a cortical microcircuit requires measurements of the spiking activity of a sufficient fraction of its neurons simultaneously. Two-photon imaging of somatic calcium signals has emerged as the method of choice thanks to its ability to record from a large number of closely spaced cells deep in intact tissues (Stosiek et al., 2003; Ko et al., 2011, 2013; Hofer et al., 2011). Recent advances have made it possible to simultaneously record from nearly every cell in a three-dimensional volume of tissue *in vivo* (Reddy and Saggau, 2005; Katona et al., 2012; Cotton et al., 2013). However, somatic calcium signals are imperfect approximations of the spiking activity: their low temporal resolution, optical contamination, and motion artifacts impose limitations on the inference of functional connectivity compared to electrophysiological recordings (Gobel and Helmchen, 2007; Grewe et al., 2010; Cotton et al., 2013).

We used the technique developed by Cotton et al. (2013) for *in vivo* high-speed 3D two-photon imaging of calcium signals to record the spiking activity of nearly every cell in volumes 200 μm wide by 100 μm deep in the superficial layers — layer 2/3 and the top portion of layer 4 — of mouse primary visual cortex. Although two-photon imaging does not allow recording from deeper layers, the tangential dimensions of the recorded cluster match the spatial extent of the dense synaptic connectivity (Song et al., 2005; Fino and Yuste, 2011; Packer and Yuste, 2011; Perin et al., 2011) that may delineate the putative cortical microcircuits. Therefore, our recordings may already begin to approach

the population sizes and densities sufficient to capture the majority of direct interactions in the center of the cluster and, perhaps, to get a glimpse of the overall functional organization of the superficial part of the cortical microcircuit.

1.3 Measures of functional connectivity

Empirical studies of multineuronal activity pursue two distinct but related goals: One is to characterize the coding of information by the population (Zohary et al., 1994; Averbeck et al., 2006; Ecker et al., 2011; Pillow et al., 2011). The other is to characterize the physiological interactions between neurons that explain their joint activity (Gerstein and Perkel, 1969; Feldt et al., 2011; Denman and Contreras, 2013). The two goals are related since interactions between neurons induce statistical dependence between their spike trains, which in turn affects the joint information they carry. However, the two goals may require different optimizations of statistical descriptions.

I define *functional connectivity* as a statistical description of multineuronal activity used for characterizing interactions between the elements of the neural circuit. For alternative definitions of this and related terms, see Feldt et al. (2011); Friston (2011). The ultimate goal of functional connectivity is to relate function to structure: to describe how interactions between cells are organized with respect to their physical arrangement, cell type, and synaptic connectivity.

The most popular measure of functional connectivity has been the Pearson correlation coefficient between the spike trains of cell pairs, often in the form of a crosscorrelogram (Gerstein and Clark, 1964; Gerstein and Perkel, 1969; Smith and Kohn, 2008; Denman and Contreras, 2013; Smith et al., 2013; Smith and Sommer, 2013; Sadovsky and MacLean, 2014). Another general approach identifies subsets of co-activated neurons called *cell assemblies* or *ensembles* as the organizing principle of multineuronal ac-

tivity (Gerstein et al., 1989; Kenet et al., 2003; Harris, 2005; Ch’Ng and Reid, 2010; Miller et al., 2014). Recently, studies of population activity have also made use of probabilistic models that model pairwise interactions as well as coactivations (Stevenson et al., 2008; Pillow et al., 2008; Hertz et al., 2011; Ganmor et al., 2011; Köster et al., 2013; Tkačík et al., 2013).

In two-photon calcium recordings, pairwise correlations remain in common use as expressions of pairwise interactions between neurons often interpreted to signify anatomical connectivity (Golshani et al., 2009; Hofer et al., 2011; Malmersjö et al., 2013; Sadovsky and MacLean, 2014).

1.4 Main contribution: Functional connectivity through partial correlations

Correlations between pairs of neurons are *marginal* properties of their activity calculated without controlling for the activity of other cells. As such, correlations may express associations arising through interactions with other cells. In multineuronal recordings pairwise correlations could be improved as a measure functional connectivity by conditioning on the activity of the rest of the circuit. In general such conditioning requires assuming some model of interactions between neurons. However, when interactions are modeled linearly, such conditioning takes the form of *partial correlations*, *i.e.* correlations that remain after removing correlations due to other variables (Whittaker, 1990). For example, partial correlations with respect to the stimulus are referred to as *noise correlations* and are used in studies of functional connectivity. Furthermore, Ecker et al. (2014) found that partial correlation with respect to common activity of the entire population could provide more consistent measurements of noise correlations across differ-

ent brain states. This study suggests that *explaining away* components of correlations that originate from outside the local circuit could reveal a more precise view of the local interactions.

In Yatsenko et al. (in review), we proposed measuring functional connectivity in two-photon recordings through partial correlations between cell pairs with respect to *all* the observed variables including the stimulus, the remaining recorded cells, and common fluctuations across the recorded population. This study empirically demonstrated that partial correlations yield a closer correspondence to the physical and functional aspects of the cortical circuit than conventional marginal correlations but the estimation of connectivity required empirically optimized regularization.

1.5 Overview

Chapter 2 reproduces Yatsenko et al. (in review) in its entirety including the background information, derivations, results, discussion, and methods.

Chapter 3 contains yet unpublished analysis of the structure of functional connectivity in cortical microcircuits with genetically labeled parvalbumin-positive (PV+) interneurons. Remarkably, partial correlations revealed clear differences in connectivity of PV+ cells compared to non-PV+ cells suggesting several possible physiological interpretations to be test in future experiments.

Finally, Chapter 4 addresses questions of physiological significance and outlines future research.

Chapter 2

Improved estimation and interpretation of correlations in neural circuits

This work has been submitted for publication in Yatsenko D, Josić K, Ecker AS, Froudarakis E, Cotton RJ, and Tolias AS (in review) Improved estimation and interpretation of correlations in neural circuits, *PLoS Computational Biology*.

Ambitious projects aim to record the activity of ever larger and denser neuronal populations *in vivo*. Correlations in neural activity measured in such recordings can reveal important aspects of neural circuit organization. However, estimating and interpreting large correlation matrices is statistically challenging. Estimation can be improved by regularization, *i.e.* by imposing a structure on the estimate. The amount of improvement depends on how closely the assumed structure represents dependencies in the data. Therefore, the selection of the most efficient correlation matrix estimator for a given neural circuit must be determined empirically. Importantly, the identity and structure of the most efficient estimator informs about the types of dominant dependencies governing the system. We sought statistically efficient estimators of neural correlation matrices in recordings from large, dense groups of cortical neurons. Using fast 3D random-access laser scanning microscopy of calcium signals, we recorded the activity of nearly every neuron in volumes 200 μm wide and 100 μm deep (150–350 cells) in mouse visual cortex. We hypothesized that in these densely sampled recordings, the correlation matrix should be best modeled as the combination of a sparse graph of pairwise partial correlations representing interactions between the observed neuronal pairs and a low-rank component representing common fluctuations and external inputs. Indeed, in cross-validation tests, the covariance matrix estimator with this structure consistently outperformed other regularized estimators. The sparse component of the estimate defined a graph of interactions. These interactions reflected the physical distances and orientation tuning properties of cells: The density of positive ‘excitatory’ interactions decreased rapidly with geometric distances and with differences in orientation preference whereas negative ‘inhibitory’ interactions were less selective. Because of its superior performance, this ‘sparse + latent’ estimator likely provides a more physiologically relevant representation of the functional connectivity in densely sampled recordings than the sample correlation matrix.

2.1 Introduction

Functional connectivity is a statistical description of observed *multineuronal* activity patterns not reducible to the response properties of the individual cells. Functional connectivity reflects local synaptic connections, shared inputs from other regions, and endogenous network activity. Although functional connectivity is a phenomenological description without a strict mechanistic interpretation, it can be used to generate hypotheses about the anatomical architecture of the neural circuit and to test hypotheses about the processing of information at the population level.

Pearson correlations between the spiking activity of pairs of neurons are among the most familiar measures of functional connectivity (Averbeck et al., 2006; Zohary et al., 1994; Kohn and Smith, 2005; Bair et al., 2001; Ecker et al., 2010). In particular, *noise correlations*, *i.e.* the correlations of trial-to-trial response variability between pairs of neurons, have a profound impact on stimulus coding (Zohary et al., 1994; Abbott and Dayan, 1999; Sompolinsky et al., 2001; Nirenberg and Latham, 2003; Averbeck et al., 2006; Josic et al., 2009; Berens et al., 2011; Ecker et al., 2011). In addition, noise correlations and correlations in spontaneous activity have been hypothesized to reflect aspects of synaptic connectivity (Gerstein and Clark, 1964). Interest in neural correlations has been sustained by a series of discoveries of their nontrivial relationships to various aspects of circuit organization such as the physical distances between the neurons (Smith and Kohn, 2008; Denman and Contreras, 2013), their synaptic connectivity (Ko et al., 2011), stimulus response similarity (Bair et al., 2001; Arieli et al., 1995; Chiu and Weliky, 2002; Kenet et al., 2003; Kohn and Smith, 2005; Cohen and Newsome, 2008; Cohen and Maunsell, 2009; Ecker et al., 2010; Rothschild et al., 2010; Ko et al., 2011; Smith and Sommer, 2013), cell types (Hofer et al., 2011), cortical layer specificity (Hansen et al., 2012; Smith et al., 2013), progressive changes in development and in learning (Golshani et al., 2009; Gu

et al., 2011; Ko et al., 2013), changes due to sensory stimulation and global brain states (Greenberg et al., 2008; Goard and Dan, 2009; Kohn et al., 2009; Rothschild et al., 2010; Ecker et al., 2014; Renart et al., 2010).

Neural correlations do not come with ready or unambiguous mechanistic interpretations. They can arise from monosynaptic or polysynaptic interactions, common or correlated inputs, oscillations, top-down modulation, and background network fluctuations, and other mechanisms (Perkel et al., 1967; Moore et al., 1970; Shadlen and Newsome, 1998; Salinas and Sejnowski, 2001; Ostožić et al., 2009; Rosenbaum and Josić, 2011). But multineuronal recordings do provide more information than an equivalent number of separately recorded pairs of cells. For example, the eigenvalue decomposition of the covariance matrix expresses shared correlated activity components across the population; common fluctuations of population activity may be accurately represented by only a few eigenvectors that affect all correlation coefficients. On the other hand, a correlation matrix can be specified using the *partial correlations* between pairs of the recorded neurons. The partial correlation coefficient between two neurons reflects their linear association conditioned on the activity of all the other recorded cells (Whittaker, 1990). Under some assumptions, partial correlations measure conditional independence between variables and may more directly approximate causal effects between components of complex systems than correlations (Whittaker, 1990). For this reason, partial correlations have been used to describe interactions between genes in functional genomics (Schäfer et al., 2005; Peng et al., 2009) and between brain regions in imaging studies (Varoquaux et al., 2012; Ryali et al., 2012). These opportunities have not yet been explored in neurophysiological studies where most analyses have only considered the distributions of pairwise correlations (Zohary et al., 1994; Bair et al., 2001; Smith and Kohn, 2008; Ecker et al., 2010).

However, estimation of correlation matrices from large populations presents a num-

ber of numerical challenges. The amount of recorded data grows only linearly with population size whereas the number of estimated coefficients increases quadratically. This mismatch leads to an increase in spurious correlations, overestimation of common activity (*i.e.* overestimation of the largest eigenvalues) (Ledoit and Wolf, 2004), and poorly conditioned partial correlations (Schäfer et al., 2005). The *sample correlation matrix* is an unbiased estimate of the true correlations but its many free parameters make it sensitive to sampling noise. As a result, on average, the sample correlation matrix is farther from the true correlation matrix than some structured estimates.

Estimation can be improved through *regularization*, the technique of deliberately imposing a structure on an estimate in order to reduce its estimation error (Schäfer et al., 2005; Bickel et al., 2006). To ‘impose a structure’ on an estimate means to bias (‘shrink’) it toward a reduced representation with fewer free parameters, the *target estimate*. The optimal target estimate and the optimal amount of shrinkage can be obtained from the data sample either analytically (Ledoit and Wolf, 2003, 2004; Schäfer et al., 2005) or by cross-validation (Friedman, 1989). An estimator that produces estimates that are, on average, closer to the truth for a given sample size is said to be more *efficient* than other estimators.

Although regularized covariance matrix estimation is commonplace in finance (Ledoit and Wolf, 2003), functional genomics (Schäfer et al., 2005), and brain imaging (Ryali et al., 2012), surprisingly little work has been done to identify optimal regularization of neural correlation matrices.

Improved estimation of the correlation matrix is beneficial in itself. For example, improved estimates can be used to optimize decoding of the population activity (Friedman, 1989; Berens et al., 2012). But reduced estimation error is not the only benefit of regularization. Finding the most efficient among many regularized estimators leads to insights about the system itself: the structure of the most efficient estimator is a parsi-

monious representation of the regularities in the data.

The advantages due to regularization increase with the size of the recorded population. With the advent of big neural data (Alivisatos et al., 2013), the search for optimal regularization schemes will become increasingly relevant in any model of population activity. Since optimal regularization schemes are specific to systems under investigation, the inference of functional connectivity in large-scale neural data will entail the search for optimal regularization schemes. Such schemes may involve combinations of heuristic rules and numerical techniques specially designed for given types of neural circuits.

What structures of correlation matrices best describe the multineuronal activity in specific circuits and in specific brain states? More specifically, are correlations in the visual cortex during visual stimulation best explained by common fluctuations or by local interactions within the recorded microcircuit?

To address these questions, we evaluated four regularized covariance matrix estimators that imposed different structures on the estimate. The estimators are designated as follows:

C_{sample} – sample covariance matrix, the unbiased, unregularized estimator.

C_{diag} – linear shrinkage of covariances toward zero, *i.e.* toward a diagonal covariance matrix.

C_{factor} – a low-rank approximation of the sample covariance matrix, representing inputs from unobserved shared factors (latent units).

C_{sparse} – sparse partial correlations, *i.e.* a large fraction of the *partial* correlations between pairs of neurons are set to zero.

$C_{\text{sparse+latent}}$ – sparse partial correlations between the recorded neurons *and* linear interactions with a number of latent units.

First, we used simulated data to demonstrate that the selection of the optimal estimator indeed pointed to the true structure of the dependencies in the data.

We then performed a cross-validated evaluation to establish which of the four regularized estimators was most efficient for representing the population activity of dense groups of neurons in mouse primary visual cortex recorded with high-speed 3D random-access two-photon imaging of calcium signals. In our data, the sample correlation coefficients were largely positive and low. We found that the best estimate of the correlation matrix was $C_{\text{sparse+latent}}$. This estimator revealed a sparse network of partial correlations ('interactions'), between the observed neurons; it also inferred latent units exerting linear effects on the observed neurons. We analyzed these networks of partial correlations and found the following: Whereas significant noise correlations were predominantly positive, the inferred interactions had a large fraction of negative values possibly reflecting inhibitory circuitry. Moreover, we found that these interactions exhibited a stronger relationship to the physical distances and to the differences in preferred orientations than noise correlations. In contrast, the inferred negative interactions were less selective.

Results

Covariance estimation

The covariance matrix is defined as

$$\Sigma = \mathbb{E} \left[(x - \mu)(x - \mu)^T \right], \quad \mu = \mathbb{E}[x] \quad (1)$$

where $\mathbb{E}[\cdot]$ denotes expectation; the $p \times 1$ vector x is a single observation of the firing rates of p neurons in a time bin of some duration; and μ is the vector of expected firing rates.

Given a set of observations $\{x(t) : t \in T\}$ of population activity, where $x(t)$ is a $p \times 1$ vector of firing rates in time bin t , and an unbiased estimate \bar{x} of the mean activity, the *sample covariance matrix* is

$$C_{\text{sample}} = \frac{1}{n} \sum_{t \in T} (x(t) - \bar{x})(x(t) - \bar{x})^T, \quad (2)$$

where n is the number of time bins in T . The sample covariance matrix is an unbiased estimate of the covariance matrix ($\mathbb{E}[C_{\text{sample}}] = \Sigma$) provided that the estimate of the mean, \bar{x} , is known or is already estimated from an independent dataset. When the mean is estimated from the same sample, C_{sample} becomes biased toward zero. However, in all the cases where unbiasedness is required, we will estimate the mean from a separate sample.

Given any covariance matrix estimate C , the corresponding correlation matrix R is calculated by normalizing the rows and columns of C by the square roots of its diagonal elements to produce unit entries on the diagonal:

$$R = (\text{diag}(C))^{-\frac{1}{2}} C (\text{diag}(C))^{-\frac{1}{2}}, \quad (3)$$

where $\text{diag}(C)$ denotes the diagonal matrix with the diagonal elements from C .

The *partial correlation* between a pair of variables is the Pearson correlation coefficient of the residuals of the linear least-squares predictor of their activity based on all the other variables, excluding the pair (Anderson, 2003; Whittaker, 1990). Partial correlations figure prominently in probabilistic *graphical modeling* wherein the joint distribu-

tion is explained by sets of two-way interactions (Whittaker, 1990). For the multivariate Gaussian distribution, zero partial correlations indicate conditional independence of the pair, implying a lack of direct interaction (Dempster, 1972; Whittaker, 1990). More generally, partial correlations can serve as a measure of conditional independence under the assumption that most dependencies in the system include strong linear effects (Whittaker, 1990; Baba et al., 2004). As neural recordings become increasingly dense, partial correlations may prove useful as indicators of conditional independence (lack of functional connectivity) between pairs of neurons.

Pairwise partial correlations are closely related to the elements of the *precision matrix*, *i.e.* the inverse of the covariance matrix (Dempster, 1972; Whittaker, 1990). Zero elements in the precision matrix signify zero partial correlation between the two variables. Given the covariance estimate C , the matrix of partial correlations P is computed by normalizing the rows and columns of the *precision matrix* C^{-1} to produce negative unit entries on the diagonal:

$$P = -(\text{diag}(C^{-1}))^{-\frac{1}{2}} C^{-1} (\text{diag}(C^{-1}))^{-\frac{1}{2}} \quad (4)$$

Increasing the number of recorded neurons results in a higher *condition number* of the covariance matrix Ledoit and Wolf (2004) making the partial correlation estimates *ill-conditioned*: small errors in the covariance estimates translate into much greater errors in the estimates of partial correlations. With massively multineuronal recordings, partial correlations cannot be estimated without *regularization* Ledoit and Wolf (2004); Schäfer et al. (2005).

We considered four regularized estimators based on distinct families of target estimates: C_{diag} , C_{factor} , C_{sparse} , and $C_{\text{sparse+latent}}$. In probabilistic models with exclusively linear dependencies, the target estimates of these estimators correspond to distinct fam-

ilies of graphical models (Fig. 1 Row 1).

The target estimate of estimator C_{diag} is the diagonal matrix D containing estimates of neurons' variances. Regularization is achieved by linear *shrinkage* of the sample covariance matrix toward D , controlled by the scalar *shrinkage intensity* parameter $\lambda \in [0, 1]$:

$$C_{\text{diag}} = (1 - \lambda)C_{\text{sample}} + \lambda D \quad (5)$$

The structure imposed by C_{diag} favors (performs better with) populations with no linear associations between the neurons (Fig. 1 Row 1, A). If sample correlations are largely spurious, C_{diag} is expected to be more efficient than other estimators.

Estimator C_{factor} approximates the covariance matrix by the factor model,

$$C_{\text{factor}} = L + D, \quad (6)$$

where L is a $p \times p$ symmetric positive semidefinite matrix with low rank and D is a diagonal matrix. This approximation is at the basis of *factor analysis* (Anderson, 2003). In factor analysis, matrix L represents covariances arising from latent factors with the rank of L indicating the number of such latent variables. Matrix D contains the variances of the independent noise in the observed variables. The estimator is regularized by the selection of the rank of L . The structure imposed by C_{factor} favors conditions in which the population activity is linearly driven by a number of latent factors that affect many cells while direct interactions between the recorded cells are insignificant (Fig. 1 Row 1, B).

Estimator C_{sparse} is produced by approximating the precision matrix by a sparse matrix S :

$$C_{\text{sparse}} = S^{-1}. \quad (7)$$

where S is a sparse matrix, *i.e.* many of its off-diagonal elements are forced to zero. The

estimator is regularized by adjusting the sparsity (fraction of off-diagonal zeros) in S . Zeros in the precision matrix coincide with zero partial correlations between the same pairs of variables. Under the assumption of linearity of interactions in the system, zero partial correlation indicates conditional independence or lack of direct interaction between the pair of variables. Therefore, the structure imposed by C_{sparse} favors conditions in which neural correlations arise from direct linear effects ('interactions') between some pairs of the observed neurons (Fig. 1 Row 1, C). The problem of finding the optimal set of non-zero elements of the precision matrix is known as *covariance selection* (Dempster, 1972).

Estimator $C_{\text{sparse+latent}}$ is obtained by approximating the precision matrix as the sum of a sparse component and a low-rank component:

$$C_{\text{sparse+latent}} = (S - L)^{-1}, \quad (8)$$

where S is a sparse matrix and L is a low-rank symmetric positive semidefinite matrix. The estimator is regularized by adjusting the sparsity of S (sparsity of local interactions) and the rank of L (number of latent units) See Methods for a more detailed explanation. The structure imposed by $C_{\text{sparse+latent}}$ favors conditions in which the activity of neurons is determined by linear effects between some observed pairs of neurons and some from several latent units (Fig. 1 Row 1, D) (Chandrasekaran et al., 2010; Ma et al., 2013).

2.1.1 Simulation

We next demonstrate how the most efficient among different regularized estimators can reveal the structure of correlations. We constructed four families of 50×50 covariance matrices, each with structure that matched one of the four regularized estimators (Fig. 1 Row 2, A–D and Methods). We used these covariance matrices as the ground truth in

Figure 1: Regularized correlations matrix estimators. **Regularized estimators whose structure matches the true structure in the data are more efficient.** **Row 1.** Graphical models of the target estimates of the four respective regularized covariance matrix estimators. Recorded neurons are represented by the green spheres and latent units by the lightly shaded spheres. Edges represent conditional dependencies, *i.e.* ‘interactions’. **Row 1, A.** For estimator C_{diag} , the target estimate is a diagonal matrix, which describes systems that lack linear dependencies. **Row 1, B.** For estimator C_{factor} , the target estimate is a factor model (low-rank matrix plus a diagonal matrix), representing systems in which correlations arise due to common input from latent units. **Row 1, C.** For estimator C_{sparse} , the covariance matrix is approximated as the inverse of a sparse matrix. This approximation describes systems in which correlations arise from a sparse set of linear associations between the observed units. **Row 1, D.** For estimator $C_{\text{sparse+latent}}$, the covariance matrix is approximated as the inverse of the sum of a sparse matrix and a low-rank matrix. This approximation describes a model wherein correlations arise due to sparse associations between the recorded cells *and* due to several latent units.

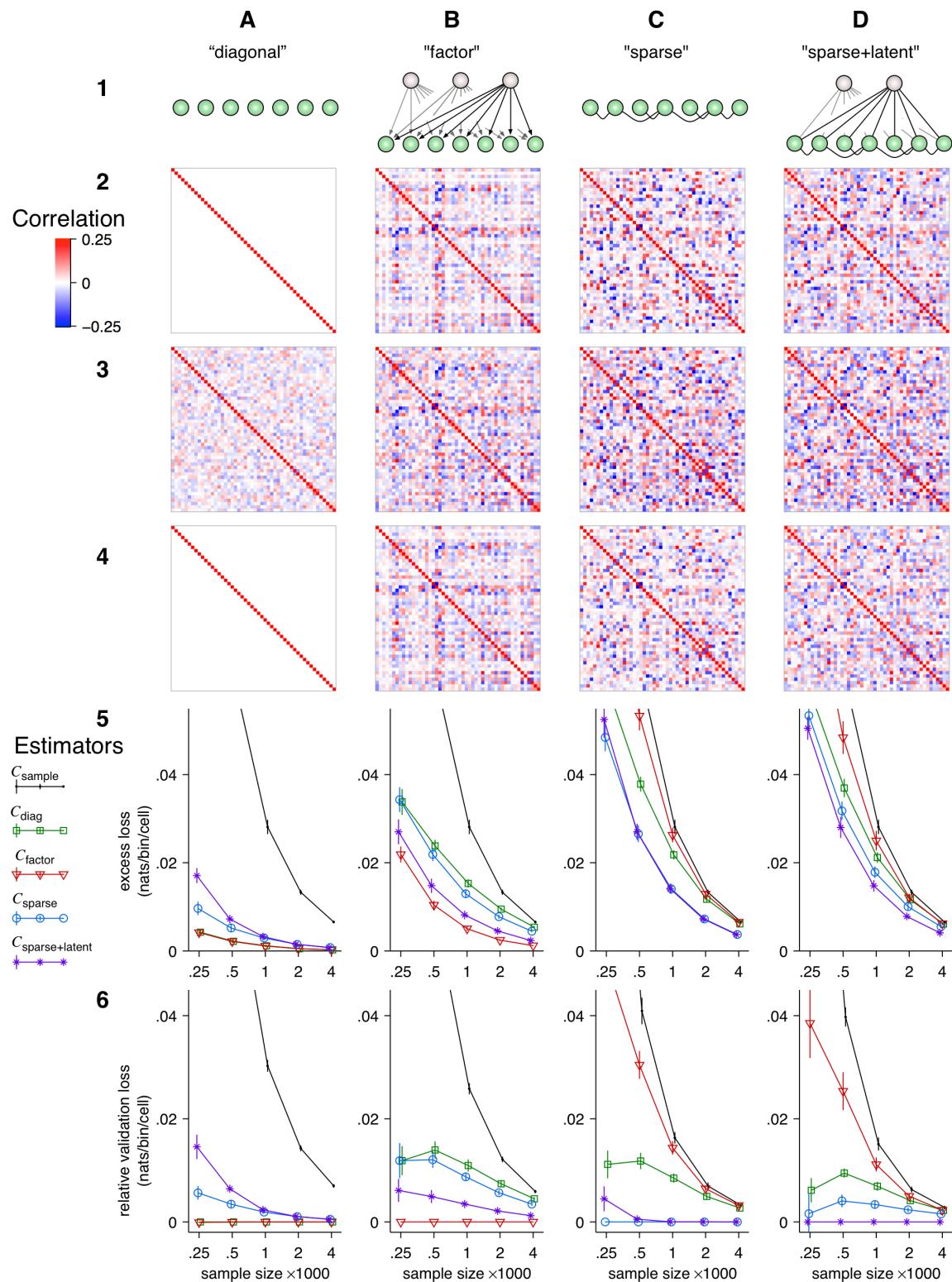
Row 2: Examples of 50×50 correlation matrices corresponding to each structure: **A.** the diagonal correlation matrix, **B.** a factor model with four latent units, **C.** a correlation matrix with 67% off-diagonal zeros in its inverse, and **D.** a correlation matrix whose inverse is the sum of a rank-3 matrix (*i.e.* three latent units) and a sparse matrix with 76% off-diagonal zeros.

Row 3: Sample correlation matrices calculated from samples of size $n = 500$ drawn from simulated random processes with respective correlation matrices shown in Row 2. The structure of the sample correlation matrix is difficult to discern by eye.

Row 4: Estimates computed from the same data as in Row 3 using structured estimators of the matching type: C_{diag} when the ground truth was diagonal in column A, C_{factor} when the ground truth is a factor model in column B, C_{sparse} when the ground truth has a sparse precision matrix in column C, and $C_{\text{sparse+latent}}$ when the ground truth has both latent units and sparse linear interactions between observed units in column C. Regularization was optimized by cross-validation. The regularized estimates are closer to the truth than the sample correlation matrices.

Row 5: True loss (Eq. 9) for the five estimators as a function of sample size. The error bars indicate the standard deviation of the mean. Estimators with structure that matches the true model converged to zero faster than the other estimators.

Row 6: Validation loss (Eq. 10) for the five estimators relative to the matching estimators for each type of ground truth. Error bars indicate the standard deviation of the mean. Differences in validation loss approximate differences in true loss.



multivariate Gaussian distributions with zero means and drew samples of various sizes. The sample correlation matrices from finite samples (*e.g.* $n = 500$ in Fig. 1 Row 3) were contaminated with sampling noise and their underlying structures were difficult to discern.

The evaluation of any covariance matrix estimator, C , is performed with respect to a *loss function* $\ell(C, \Sigma)$ to quantify its discrepancy from the truth, Σ . The loss function is chosen to attain its minimum when $C = \Sigma$. Here, in the role of the loss function we adopted the Kullback-Leibler divergence between multivariate normal distributions with equal means, scaled by $\frac{1}{p}$ to make its values comparable across different population sizes:

$$\ell(C, \Sigma) = \frac{1}{p} D_{KL}(\mathcal{N}_\Sigma \| \mathcal{N}_C) = \frac{1}{2p} [\text{tr}(C^{-1}\Sigma) + \ln \det C - \ln \det \Sigma - p] \quad (9)$$

Therefore, $\ell(C, \Sigma)$ is expressed in nats/neuron per time bin.

We drew 30 independent samples with sample sizes $n = 250, 500, 1000, 2000$, and 4000 from each model and computed the loss $\ell(C, \Sigma)$ for each of the five estimators. The hyperparameters of the regularized estimators were optimized by nested cross-validation using only the data in the sample. All the regularized estimators produced better estimates (lower loss) than the sample covariance matrix. However, estimators whose structure matched the true model outperformed the other estimators (Fig. 1 Row 5).

Note that when the ground truth had zero correlations (Column A), C_{factor} performed equally well to C_{diag} because it correctly inferred zero factors and only estimated the individual variances. Similarly, when the number of latent units was zero (Column C), $C_{\text{sparse+latent}}$ performed nearly equally well to C_{sparse} because it correctly inferred zero latent units. With increasing sample sizes, all estimators converged to the ground truth (zero loss) but the estimators with correct structure outperformed the others even for

large samples.

In more realistic conditions, when the ground truth is not accessible, the loss cannot be computed directly but may be estimated from data through *validation*. In a validation procedure, a validation sample covariance matrix C'_{sample} is computed from a testing data set that is independent from the data used for computing C . Then the *validation loss* $\mathcal{L}(C, C'_{\text{sample}})$ measures the discrepancy of C from C'_{sample} . Here, in the role of validation loss, we adopted the negative multivariate normal log likelihood of C given C'_{sample} , also scaled by $\frac{1}{p}$ and omitting the constant term:

$$\mathcal{L}(C, C'_{\text{sample}}) = \frac{1}{2p} \left[\text{tr}(C^{-1} C'_{\text{sample}}) + \ln \det C \right] \quad (10)$$

Since $\mathcal{L}(\cdot, \cdot)$ is additive in its second argument and C'_{sample} is an unbiased estimate of Σ , then, for given C and Σ , the validation loss is an unbiased estimate of the true loss, up to a constant:

$$\mathbb{E} \left[\mathcal{L}(C, C'_{\text{sample}}) \right] = \mathcal{L} \left(C, \mathbb{E} \left[C'_{\text{sample}} \right] \right) = \mathcal{L}(C, \Sigma) = \ell(C, \Sigma) + \text{const.} \quad (11)$$

Therefore, the validation procedure allows comparing the relative values of loss of different covariance estimators.

Indeed, the validation loss computed by 10-fold cross-validation (see Methods) accurately reproduced the relative values of the true loss and the rankings of the covariance estimators without access to the ground truth (Fig. 1 Row 6).

In the example above, the data were sampled from multivariate normal random variables. In such models, partial correlations perfectly characterize the conditional dependencies between variables and the graphical models of partial correlations exactly correspond to the conditional dependencies in the data. To demonstrate that estimator

rankings were robust to deviations from Gaussian models, we repeated the same cross-validated evaluation using pairwise Ising models to generate the data.

Ising models have been used to infer functional connectivity from neuronal spike trains (Hertz et al., 2011). Conveniently, the Ising model has equivalent mathematical form to the Gaussian distribution,

$$x \sim \frac{1}{Z(J, h)} \exp\left(\frac{1}{2} x^\top J x + h^\top x\right) \quad (12)$$

but the Ising model is defined on the multivariate binary domain rather than the continuous domain. Both models are maximum-entropy models constrained to match the mean firing rates and the covariance matrix (Jaynes, 1957). The partition function $Z(J, h)$ normalizes the distribution on the models' respective domains. In the Gaussian model, the matrix $-J^{-1}$ is the covariance matrix; and the mean values are $\mu = J^{-1}h$. For the Ising model, J is the matrix of pairwise interactions and h is the vector of the cells' individual activity drives, although they do not have a simple relationship to the means and the covariance matrix. Both distributions have the same structure of pairwise conditional dependencies: zeros in the matrix J indicate conditional independence between the corresponding pair of neurons.

Indeed, despite their considerable departure from strictly linear conditional dependencies, Ising models yielded the same relationships between the performances of the covariance estimators as the Gaussian models in cross-validation (Fig. 2). Identical interaction matrices J of the joint distributions over the observable and latent variables were used for both the Gaussian and the Ising models.

This simulation study demonstrated that cross-validated evaluation of regularized estimators of the covariance matrix of population activity can discriminate between structures of dependencies in the population. The identity of the most efficient covari-

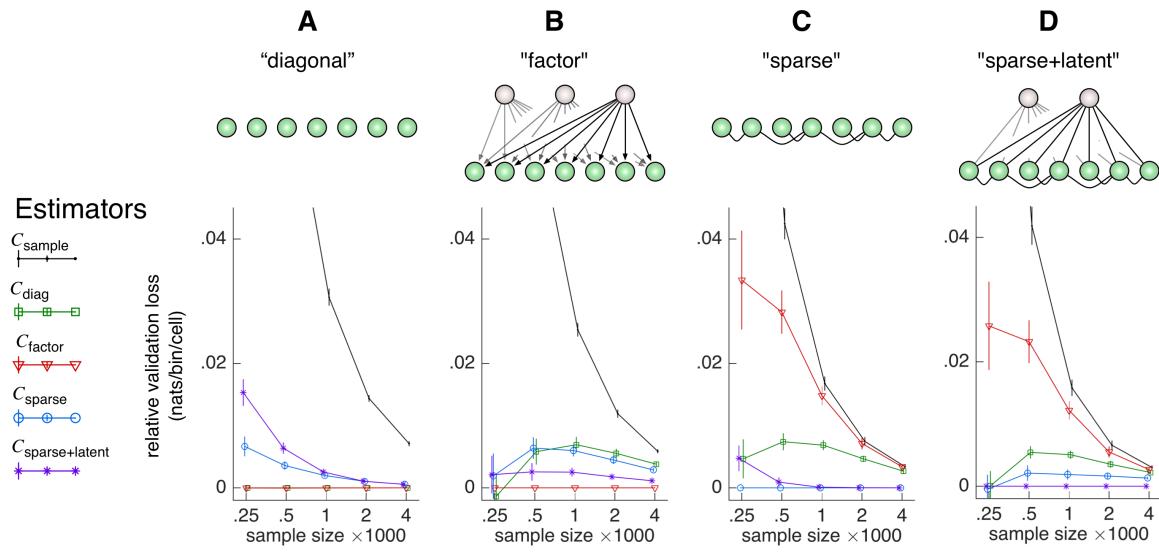


Figure 2: Performance of covariance estimators on samples from Ising models. A–D

Validation losses of covariance matrix estimators relative to the estimator whose structure matches the ground truth. The calculation is performed identically to Fig. 1 Row 6 except Ising models are used as ground truth instead of Gaussian models. This cross-validated evaluation correctly ranked the estimators despite the models' deviation from Gaussianity.

ance estimator for a particular neural circuit is therefore an empirical finding characterizing the nature of circuit interactions.

The $C_{\text{sparse+latent}}$ estimator is most efficient in neural data

We recorded the calcium activity of densely sampled populations of neurons in layers 2/3 and upper layer 4 in primary visual cortex of sedated mice using fast random-access 3D scanning two-photon microscopy during visual stimulation (Fig. 3 A–B) (Reddy and Saggau, 2005; Katona et al., 2012; Cotton et al., 2013). This technique allowed fast sampling (100–150 Hz) from large numbers (150–350) of cells in a small volume of cortical tissue ($200 \times 200 \times 100 \mu\text{m}^3$) in layers 2/3 and 4 (Fig. 3 C and D). The firing rates were inferred using sparse nonnegative deconvolution (Vogelstein et al., 2010) (Fig. 3 C). Only cells that produced detectable calcium activity were included in the analysis (see Methods). First, 30 repetitions of full-field drifting gratings of 16 directions were presented in random order. Each grating was displayed for 500 ms, without intervening blanks. This stimulus was used to compute the orientation tuning of the recorded cells (Fig. 3 D). To estimate the noise correlation matrix, we presented only two distinct directions in some experiments or five directions in others with 100–300 repetitions of each direction. Each grating lasted 1 second and was followed by a 1-second blank. The traces were then binned into 150 ms intervals aligned on the stimulus onset for the estimation of the correlation matrix. The sample correlation coefficients were largely positive and low (Fig. 3 E and F). The average value of the correlation coefficient across sites ranged from 0.0065 to 0.051 with the mean across sites of 0.018 (Fig. 7 D).

In these densely sampled populations, direct interactions between cells are likely to influence the patterns of population activity. We therefore hypothesized that covariance matrix estimators that explicitly modeled the partial correlations between pairs

of neurons (C_{sparse} and $C_{\text{sparse+latent}}$) would have a performance advantage. However, the observed neurons must also be strongly influenced by global activity fluctuations and by unobserved common inputs to the advantage of estimators that explicitly model common fluctuations of the entire population: C_{factor} and $C_{\text{sparse+latent}}$. If both types of effects are significant, then $C_{\text{sparse+latent}}$ should outperform the other estimators.

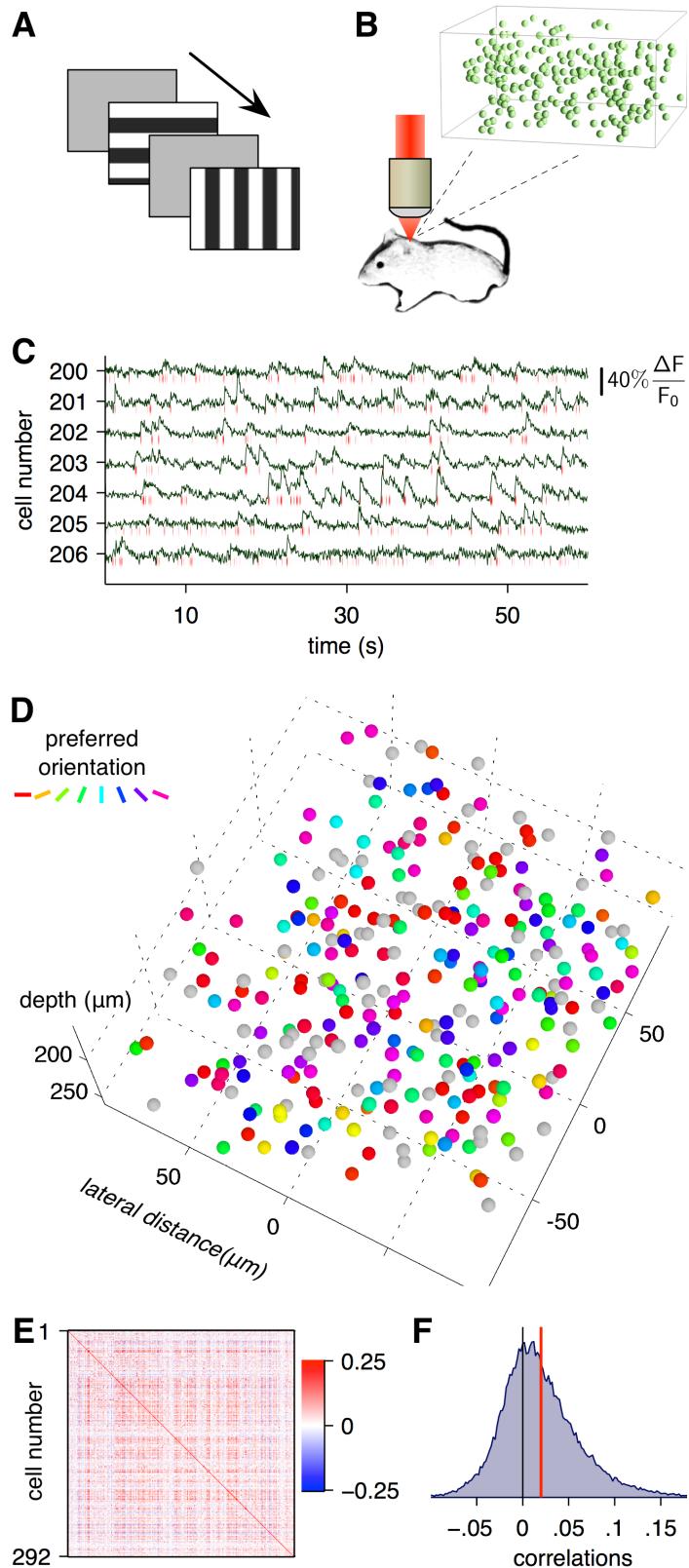
To test this hypothesis, we computed the relative validation loss of estimators C_{sample} , C_{diag} , C_{factor} , and C_{sparse} with respect to $C_{\text{sparse+latent}}$ in $n = 27$ imaged sites in 14 mice. The hyperparameters of each estimator were optimized by nested cross-validation (See Fig. 5 and Methods). Indeed, the sparse+latent estimator outperformed the other estimators (Fig. 4). The respective median differences of the validation loss were 0.039, 0.0016, 0.0029, and 0.0059 nats/cell/bin, significantly greater than zero ($p < 0.01$ in each comparison, $n = 27$ sites in 14 mice, Wilcoxon signed rank test).

Structure of $C_{\text{sparse+latent}}$ estimates

We examined the composition of the $C_{\text{sparse+latent}}$ estimates at each imaged site (Fig. 6 and Fig. 7). Although the regularized estimates were similar to the sample correlation matrix (Fig. 6 A and B), the corresponding partial correlation matrices differed substantially (Fig. 6 C and D). The estimates separated two sources of correlations: a network of linear interactions expressed by the sparse component of the inverse and latent units expressed by the low-rank components of the inverse (Fig. 6 E). The sparse partial correlations revealed a network that differed substantially from the network composed of the greatest coefficients in the sample correlation matrix (Fig. 6 F, G, H, and I).

In the example site (Fig. 6), the sparse component had 92.8% sparsity (or conversely, 7.2% connectivity: connectivity = 1 – sparsity) with average node degree of 20.9 (Fig. 6 G). The average node degree, *i.e.* the average number of interactions linking each neuron,

Figure 3: Estimation of noise correlations from population calcium signals. Visual stimuli comprising full-field drifting gratings interleaved with blank screens (**A**) presented during two-photon recordings of somatic calcium signals using fast 3D random-access microscopy (**B**). **C–F**. Calcium activity data from an example site. **C**. Representative calcium signals of seven cells, downsampled to 20 Hz, out of the 292 total recorded cells. Spiking activity inferred by nonnegative deconvolution is shown by red ticks below the trace. **D**. The spatial arrangement and orientation tuning of the 292 cells from the imaged site. The cells' colors indicate their orientation preferences. The gray cells were not significantly tuned. **E**. The sample noise correlation matrix of the activity of the neural population. **F**. Histogram of noise correlation coefficients in one site. The red line indicates the mean correlation coefficient of 0.020.



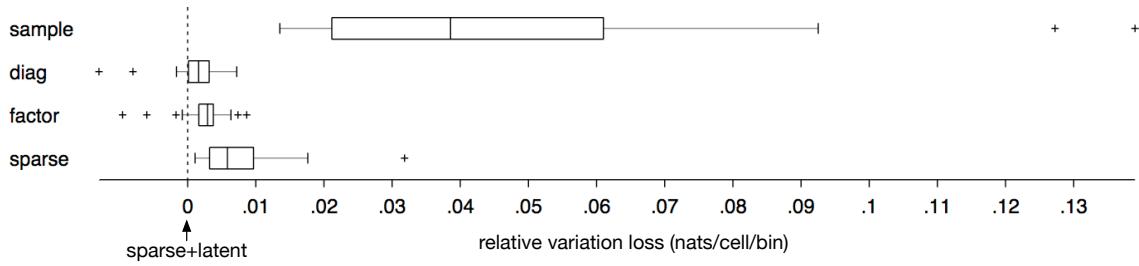


Figure 4: Evaluation of correlation estimators by cross-validation The plot depicts the validation losses (eq. 10) of estimators C_{sample} , C_{diag} , C_{factor} , and C_{sparse} minus the validation loss of $C_{\text{sparse+latent}}$ indicated by the dashed vertical line. The difference is consistently positive ($p < 0.01$ in each comparison, Wilcoxon signed rank test, $n = 27$ sites in 14 mice), indicating that $C_{\text{sparse+latent}}$ consistently outperforms the other estimators in these neural data. The box plots indicate the 25th, 50th, and 75th percentiles with the whiskers extending to the minimum and maximum values after excluding the outliers marked with '+'.

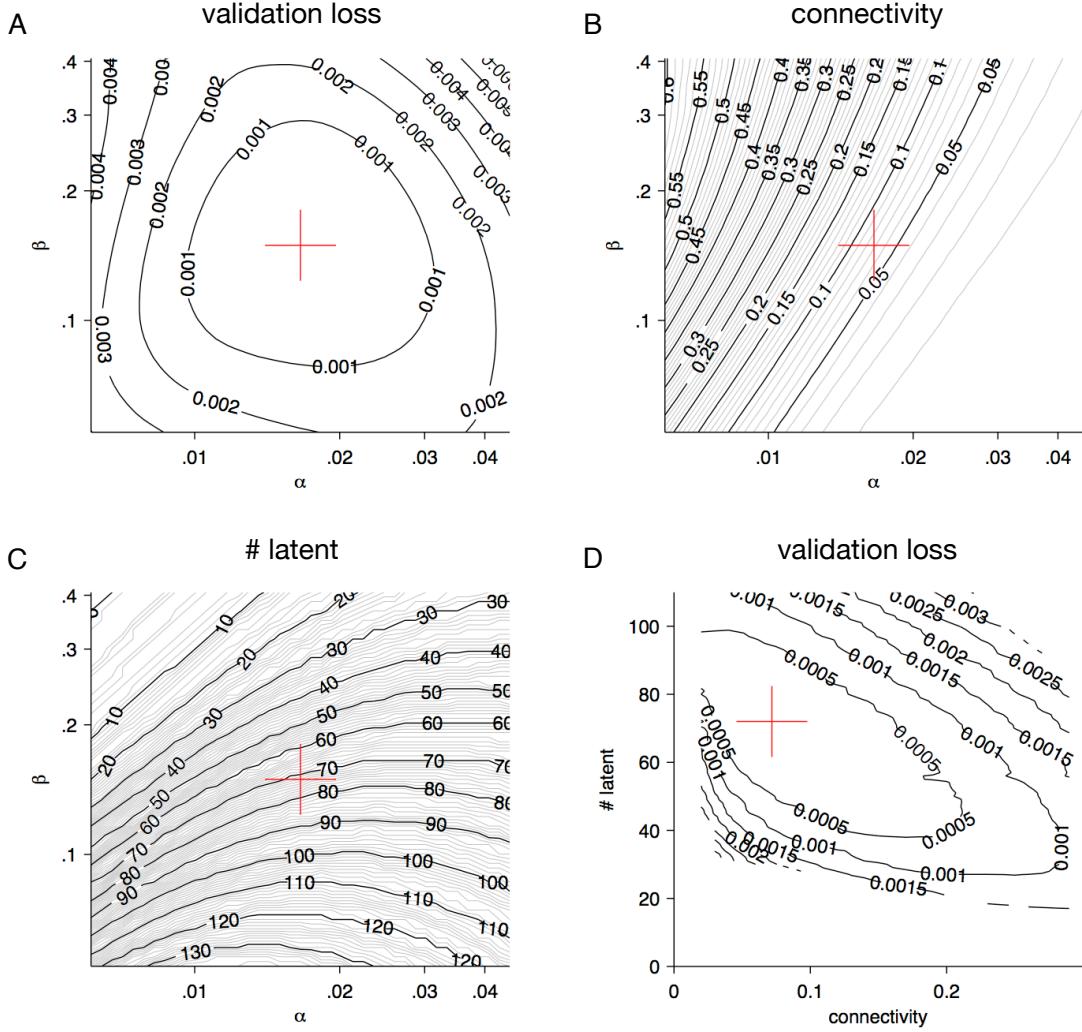
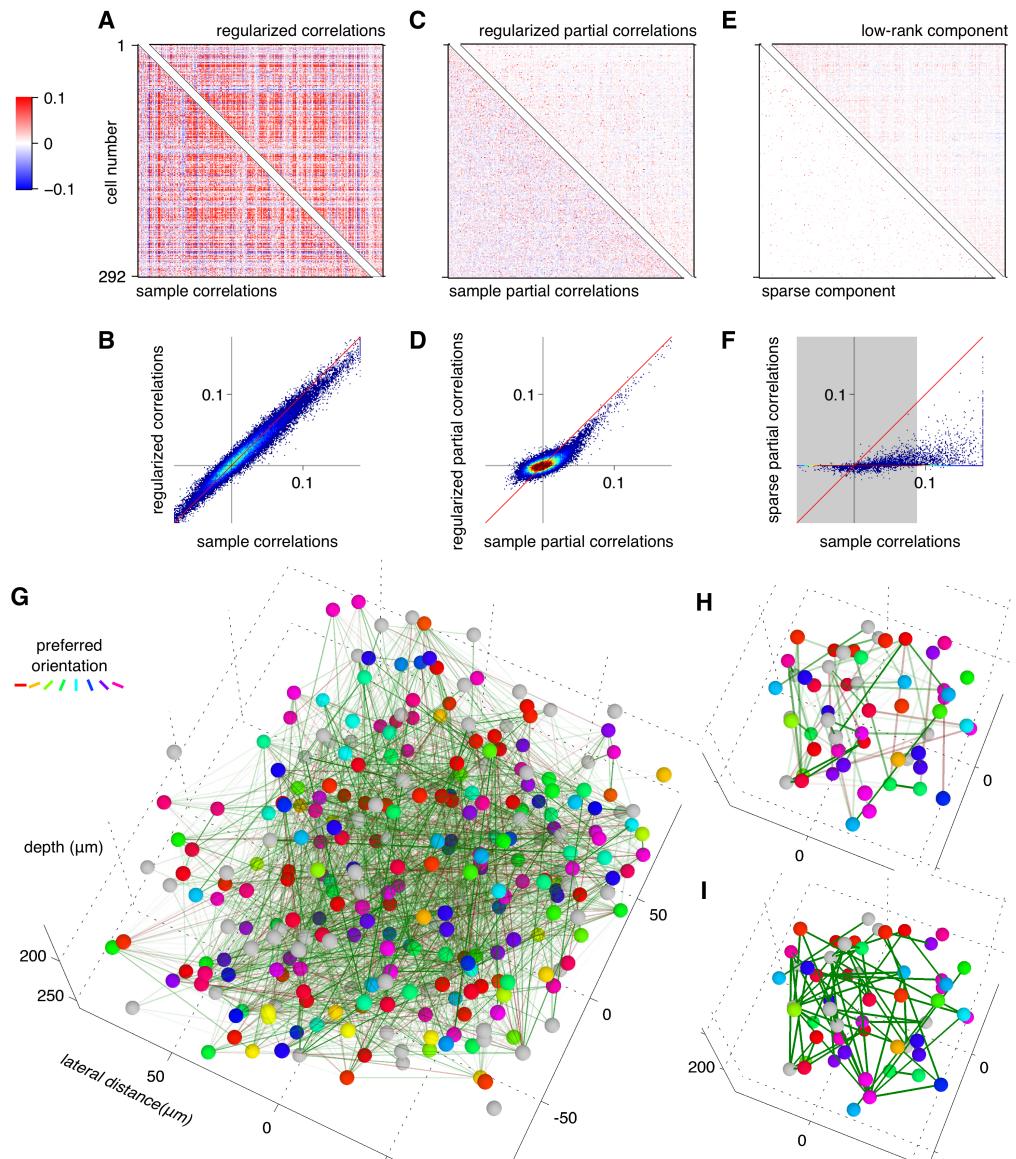


Figure 5: Optimization of regularization parameters of the $C_{\text{sparse+latent}}$ estimator. A. Validation loss (Eq. 25) for the example site in Fig. 2 and 4 as a function of the hyperparameters α and β of the $C_{\text{sparse+latent}}$ estimator (Eq. 8 and Eq. 18). In all panels, the red cross marks the optimal value found by the pattern search algorithm described in Methods. **B.** The connectivity (1 – sparsity) of the sparse component S as a function of α and β for the example site. **C.** The number of latent units, *i.e.* the rank of the low-rank component L , as a function of hyperparameters α and β . **D.** The loss function as a function of the connectivity and the number of latent units.

Figure 6: Example of sparse partial correlation structure revealed by $C_{\text{sparse+latent}}$. **A, B.** The regularized estimate $C_{\text{sparse+latent}}$ closely approximates the sample correlation matrix C_{sample} . **C, D.** The partial correlation matrices from the two estimates differ substantially. **E.** The partial correlation matrix of the regularized estimate is decomposed into a sparse component with 92.8% off-diagonal zeros (bottom-left) and low-rank component of rank 72 (top-right). **F.** The sparse component of the regularized partial correlation matrix had little resemblance to the sample correlations: The gray region indicates the range of correlations containing 92.8% of cells pairs, equal to the fraction of zeros in the sparse partial correlation matrix. Correlation coefficients outside this interval formed the network of greatest correlations. This network differed from the sparse component of the $C_{\text{sparse+latent}}$: Only 27.7% of the highest correlations coefficients outside the gray regions coincided with interactions inferred by $C_{\text{sparse+latent}}$. **G.** A graphical depiction of the positive (green) and negative (magenta) sparse partial correlations as edges between observed neurons. The line density is proportional to the magnitude of the partial correlation. **H.** A subset of neurons from the center of the cluster shown in **G** showing the sparse partial correlations. **I.** The same subset of neurons with edges indicating sample correlations thresholded to match the sparsity of the sparse partial correlation. These edges correspond to the sample correlation coefficients outside the gray region in panel F.



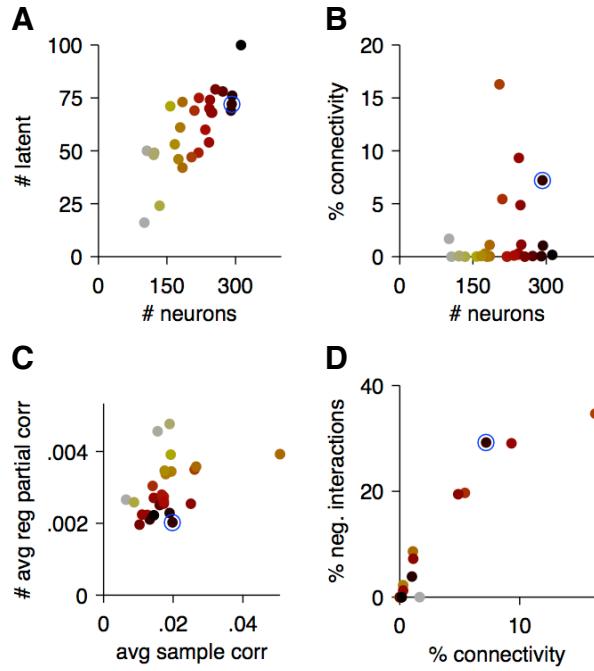


Figure 7: Properties of $C_{\text{sparse+latent}}$ estimates from all imaged sites. Each point represents an imaged site with its color indicating the population size as shown in panels A and B. The example site from Figures 3 and 6 is circled in blue.

A. The number of inferred latent units *vs.* population size. **B.** The connectivity of the sparse component of partial correlations as a function of population size. **C.** The average sample correlations *vs.* the average partial correlations (Eq. 4) of the $C_{\text{sparse+latent}}$ estimate. **D.** The percentage of negative interactions *vs.* connectivity in the $C_{\text{sparse+latent}}$ estimates.

is related to connectivity as $\text{degree} = \text{connectivity} \cdot (p - 1)$, where p is the number of neurons. The low-rank component had rank 72, denoting 72 inferred latent units. The number of latent units increased with population size (Fig. 7 A) but the connectivity was highly variable (Fig. 7 B): Several sites, despite their large population sizes, were driven by latent units and had few pairwise interactions. This variability may be explained by differences in brain states and recording quality and warrants further investigation.

The average partial correlations calculated from these estimates according to Eq. 4 at all 27 sites were about 5 times lower than the average sample correlations (Fig. 7 C). This suggests that correlations between neurons build up from multiple chains of smaller interactions. Furthermore, the average partial correlations were less variable: the coefficient of variation of the average sample correlations across sites was 0.45 whereas that of the average partial correlations was 0.29, with larger populations exhibiting greater uniformity of average partial correlations than the smaller populations ($p = 0.002$ Brown-Forsythe test).

While the sample correlations were mostly positive, the sparse component of the partial correlations ('interactions') had a high fraction (28.7% in the example site) of negative values (Fig. 6 F). The fraction of negative interactions increased with the inferred connectivity (Fig. 7 D), suggesting that negative interactions can be inferred only after a sufficient density of positive interactions has been uncovered.

Thresholded sample correlations have been used in several studies to infer pairwise interactions (Golshani et al., 2009; Feldt et al., 2011; Malmersjö et al., 2013; Sadovsky and MacLean, 2014). We therefore compared the interactions in the sparse component of $C_{\text{sparse+latent}}$ to those obtained from the sample correlations thresholded to the same level of connectivity. The networks revealed by the two methods differed substantially. In the example site with 7.2% connectivity in $C_{\text{sparse+latent}}$, only 27.7% of the connections coincided with the above-threshold sample correlations (Fig. 6 F, H, and I). In particu-

lar, most of the inferred negative interactions corresponded to low sample correlations (Fig. 6 F) where high correlations should be expected given the rest of the correlation matrix.

Relationship of $C_{\text{sparse+latent}}$ to orientation tuning and physical distances

We examined how the structure of the $C_{\text{sparse+latent}}$ estimates related to the differences in orientation preference and to the physical distances separating pairs of cells (Fig. 8). Five sites with highest pairwise connectivities were included in the analysis. Partial correlations were computed using Eq. 4 based on the regularized estimate, including both the sparse and the latent component. Connectivity was computed as the fraction of pairs of cells connected by non-zero elements (interactions) in the sparse component of the estimate, distinguishing between the positive and negative connectivities.

First, we analyzed how correlations and connectivity depended on the difference in preferred orientations (Δori) of pairs of significantly ($\alpha = 0.05$) tuned cells. The partial correlations decayed more rapidly with Δori than did sample correlations ($p < 10^{-9}$ in each of the five sites, two-sample t -test of the difference of the linear regression coefficients). Positive connectivity decreased with Δori ($p < 0.005$ in each of the five sites, t -test on the logistic regression coefficient) whereas negative connectivity did not decrease (Fig. 8 D): The slope in the logistic model of connectivity with respect to Δori was significantly higher for positive than for negative interactions ($p < 0.04$ in each of the five sites, two-sample t -test of the difference of the logistic regression coefficient).

Second, we compared how correlations and connectivity depended on the physical distance separating pairs of cells. We distinguished between lateral distance, Δx , in the plane parallel to the pia, and vertical distance, Δz , orthogonal to the pia. When considering the dependence on Δx , the analysis was limited to cell pairs located at the same

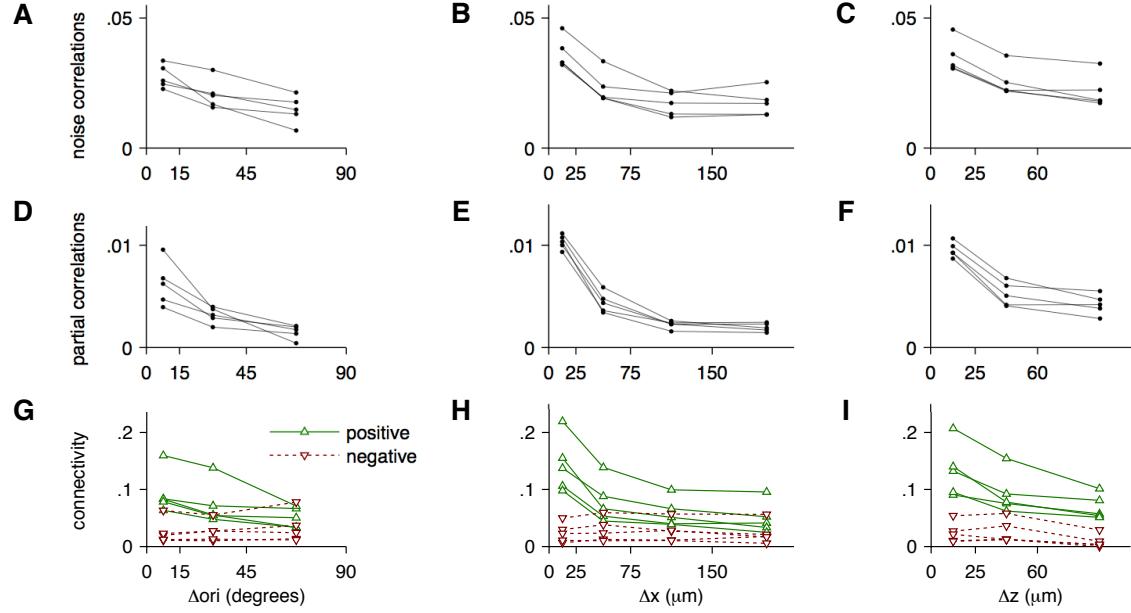


Figure 8: Relationship between functional connectivity and circuit architecture Dependence of sample correlations, regularized partial correlations, and connectivity inferred by $C_{\text{sparse+latent}}$ on the differences in preferred orientations, Δori , and physical distances: horizontal Δx and depth Δz . Five sites with highest connectivity (see Fig. 7 B) were selected for this analysis.

A–C. Mean sample correlations in relation to Δori , Δx and Δz , respectively. For Δx averages, only horizontally aligned cell pairs with $\Delta z < 30 \mu\text{m}$ were considered. Similarly, for Δz averages, only vertically aligned cell pairs with $\Delta x < 30 \mu\text{m}$ were considered.

D–F. Mean partial correlations regularized by the $C_{\text{sparse+latent}}$ estimator binned the same way as the sample correlations above. The partial correlations exhibit stronger dependence on Δori , Δx , and Δz than sample correlations.

G–I. Positive connectivity (green) and negative connectivity (red) inferred by the $C_{\text{sparse+latent}}$ estimator. Positive and negative connectivities refer to the fractions of the positive and negative partial correlations computed from the sparse component S of $C_{\text{sparse+latent}}$. Positive connectivity decreases with Δori , Δx , and Δz . Negative connectivity does not decrease with Δori , Δx within the examined range, and with Δz for small values of $\Delta z < 60 \mu\text{m}$.

depth with $\Delta z < 30 \mu\text{m}$; conversely, when considering the dependence on Δz , only vertically aligned cell pairs with $\Delta x < 30 \mu\text{m}$ were included. Again, the partial correlations decayed more rapidly both laterally and vertically than sample correlations ($p < 10^{-6}$ in each of the five sites, for both lateral and vertical distances, two-sample t -test of the difference of the linear regression coefficients). Positive connectivity decayed with distance ($p < 10^{-6}$ in each of the five sites for positive interactions and $p < 0.05$ for negative interactions, t -test on the logistic regression coefficient) (Fig. 8 E), so that cells separated laterally by less than $25 \mu\text{m}$ were 3.2 times more likely to be connected than cells separated laterally by more than $150 \mu\text{m}$. Although the positive connectivity appeared to decay faster with vertical than with lateral distance, the differences in slopes of the respective logistic regression models were not significant with available data. The negative connectivity decayed slower with distance (Fig. 8 E and F): The slope in the respective logistic models with respect to the lateral distance was significantly higher for positive than for negative connectivities ($p < 0.05$ in each of the five sites, two-sample t -test of the difference of the logistic regression coefficients).

2.2 Discussion

Functional connectivity as a network of pairwise interactions

Functional connectivity is often represented as a graph of pairwise interactions. The goal of many studies of functional connectivity has been to estimate anatomical connectivity from observed multineuronal spiking activity. For example, characteristic peaks and troughs in the pairwise cross-correlograms of recorded spike trains contain statistical signatures of directional monosynaptic connections and shared synaptic inputs (Gerstein and Clark, 1964; Perkel et al., 1967; Moore et al., 1970; Alonso and Martinez,

1998; Denman and Contreras, 2013). Such signatures are ambiguous as they can arise from network effects other than direct synaptic connections (Aertsen et al., 1989). With simultaneous recordings from more neurons, ambiguities can be resolved by inferring the conditional dependencies between pairs of neurons. Direct causal interactions between neurons produce statistical dependency between them even after conditioning on the state of the remainder of the network and external input. Therefore, conditional independence can signify the absence of a direct causal influence.

Conditional dependencies can be inferred by fitting a probabilistic model of the joint population activity. For example, generalized linear models (GLMs) have been constructed to include biophysically plausible synaptic integration, membrane kinetics, and individual neurons' stimulus drive (Pillow et al., 2008). Maximum entropy models constrained by observed pairwise correlations are among other models with pairwise coupling between cells (Schneidman et al., 2006; Tkacik et al., 2006; Yu et al., 2008; Tang et al., 2008; Shlens et al., 2009). Assuming that the population response follows a multivariate normal distribution, the conditional dependencies between pairs of neurons are expressed by the partial correlations between them. Each probabilistic model, fitted to the same data may reveal a completely different network of 'interactions', *i.e.* conditional dependencies between pairs of cells.

It is not yet clear which approach provides the best correspondence with anatomical connectivity. Little experimental evidence is available to answer this question. The connectivity graphs inferred by various statistical methods are commonly reported without examining their relation to anatomy. Topological properties of such graphs have been interpreted as principles of circuit organization (*e.g.* small-world organization) (Feldt et al., 2011; Yu et al., 2008; Malmersjö et al., 2013; Sadovsky and MacLean, 2014). However, the topological properties of functional connectivity graphs can depend on the method of inference (Zalesky et al., 2012). Until a physiological interpretation of func-

tional connectivity is established, the physiological relevance of such analyses remains in question and we did not attempt graph-theoretical analyses on inferred sparse networks of interactions.

Inference of the conditional dependencies also depends on the completeness of the recorded population: To equate conditional dependency to direct interaction between two neurons, we must record from all neurons with which the pair interacts. Unobserved portions of the circuit may manifest as conditional dependencies between observed neurons that do not interact. For this reason, statistical models of population activity have been most successfully applied to *in vitro* preparations of the retina or cell cultures where high-quality recordings from the complete populations were available (Pillow et al., 2008). In cortical tissue, electrode arrays record from a small fraction of cells in a given volume, limiting the validity of inference of the pairwise conditional dependencies. Perhaps for this reason, partial correlations have not, until now, been used to describe the functional connectivity in cortical populations.

Two-photon imaging of population calcium signals presents unique advantages for the estimation of functional connectivity. While the temporal resolution of calcium signals is limited by calcium dye kinetics, fast imaging techniques combined with spike inference algorithms provide millisecond-scale temporal resolution of single action potentials (Grewe et al., 2010). However, such high temporal precision comes at the cost of the accuracy of inferred spike rates. Better accuracy is achieved when calcium signals are analyzed on scales of tens of milliseconds (Cotton et al., 2013; Theis et al., 2014). The major advantage of calcium imaging is its ability to characterize the spatial arrangement and types of recorded cells. Recently, advanced imaging techniques have allowed recording from nearly every cell in a volume of cortical tissue *in vivo* (Katona et al., 2012; Cotton et al., 2013) and even from entire nervous systems (Leung et al., 2013; Ahrens et al., 2013). These techniques may provide more incisive measurements of functional

connectivity than electrophysiological recordings.

The low temporal resolution of calcium signals limits the use of functional connectivity methods that rely on millisecond-scale binning of signals (cross-correlograms, some GLMs, and binary maximum entropy models). Hence, most studies of functional connectivity have relied on instantaneous sample correlations (Greenberg et al., 2008; Golshani et al., 2009; Hofer et al., 2011; Malmersjö et al., 2013) . Although some investigators have interpreted such correlations as indicators of (chemical or electrical) synaptic connectivity, most used them as more general indicators of functional connectivity without relating them to underlying mechanisms.

In this study, we sought to infer pairwise functional connectivity networks in cortical microcircuits. We hypothesized that partial correlations correspond more closely to underlying mechanisms than sample correlations when recordings are sufficiently dense. Since neurons form synaptic connections mostly locally and sparsely (Perin et al., 2011), we *a priori* favored solutions with sparse partial correlations. Under the assumptions that the recorded population is sufficiently complete and that the model correctly represents the nature of interactions, the network of partial correlations can be hypothesized to be a better representation of functional dependencies than correlations.

Functional connectivity as coactivations

Another approach to describing the functional connectivity of a circuit is to isolate individual patterns of multineuronal coactivations (Gerstein et al., 1989; Chapin and Nicolelis, 1999; Peyrache et al., 2010; Ch'Ng and Reid, 2010; Lopes-dos Santos et al., 2011, 2013). Depending on the method of their extraction, coactivation patterns may be referred to as *assemblies*, *factor loadings*, *principal components*, *independent components*, *activity modes*, *eigenvectors*, or *coactivation maps*. Coactivation patterns could be interpreted

as signatures of Hebbian cell assemblies (Gerstein et al., 1989; Ch'Ng and Reid, 2010), *i.e.* groups of tightly interconnected groups of cells involved in a common computation. Coactivation patterns could also result from shared input from unobserved parts of the circuit, or global network fluctuations modulating the activity of the local circuit (Okun et al., 2012; Ecker et al., 2014).

Coactivation patterns and pairwise connectivity are not mutually exclusive since assemblies arise from patterns of synaptic connectivity. However, an analysis of coactivation shifts the focus from detailed interactions to collective behavior. In our study, the functional connectivity solely through modes of coactivations was represented by the factor analysis-based estimator C_{factor} .

Combining pairwise interactions and coactivations

In the effort to account for the joint activity patterns that are poorly explained by pairwise interactions, investigators have augmented models of pairwise interactions with additional mechanisms such as latent variables, higher-order correlations, or global network fluctuations (Ganmor et al., 2011; Tkačík et al., 2013; Pfau et al., 2013; Köster et al., 2013; Ecker et al., 2014).

In our study, we combined pairwise interactions with collective coactivations by applying the recently developed numerical techniques for the inference of the partial correlation structure in systems with latent variables (Chandrasekaran et al., 2010; Ma et al., 2013). The resulting estimator, $C_{\text{sparse+latent}}$, effectively decomposed the functional connectivity into a sparse network of pairwise interactions and coactivation mode vectors.

Addressing ill-posedness

Inferring the conditional dependencies between variables in a probabilistic model is an ill-posed problem: small variations in the data produce large errors in the inferred network of dependencies. The problem becomes worse as the number of recorded neurons increases until such models lose their statistical validity (Roudi et al., 2009). As techniques have improved to allow recording from larger neuronal populations, experimental neuroscientists have addressed this problem by extending the recording durations to keep sampling noise in check and verified that existing models are not overfitted (Tkačik et al., 2013). However, ambitious projects, such as the BRAIN initiative (Alivisatos et al., 2013), aim to record from significantly larger populations. Simply increasing recording duration will not be practical or sufficient, and the problem must be addressed by using regularized estimators. Regularization biases the solution toward a small subspace in order to counteract the effect of sampling noise in the empirical data. However, biasing the solution to an inappropriate subspace does not allow significant estimation improvement and hinders interpretation.

Several strategies have been developed to limit the model space in order to improve the quality of the estimate. For example, Ganmor et al. (Ganmor et al., 2011) developed a heuristic rule to identify the most significant features that must be fitted by a maximum entropy model for improved performance in the retina. As another example of regularization, generalized linear models typically employ L_1 penalty terms to constrain the solution space and to effectively reduce the dimensionality of the solution (Pillow et al., 2008).

In our study, regularizations were accomplished by dimensionality reduction (feature selection) schemes to produce sparse, constrained solutions. Only the most efficient scheme was considered in the analysis of functional connectivity.

Model selection

Various model selection criteria have been devised to select between families of models and the optimal subsets of variables in a given model family based on observed data. Despite its computational requirements, cross-validation is among the most popular model selection methods due to its minimal assumptions about the data generating process (Arlot and Celisse, 2010).

We evaluated the covariance matrix estimators using a loss function derived from the normal distribution. However, this does not limit the applicability of its conclusions to normal distributions. Other probabilistic models, fitted to the same data, could also serve as estimators of the covariance matrix. If a different model yields better estimation of the covariance matrix than the estimator proposed here, we believe that its structure should deserve consideration as the better representation of the functional connectivity.

The results of model selection must be interpreted with caution. As we demonstrated by simulation, even models with incorrect forms of dependencies can substantially improve estimates (Fig. 1). Therefore, showing that a more constrained model has better cross-validated performance than a more complex model does not necessarily support the conclusion that it reveals a better representation of dependencies in the data. This caveat is related to *Stein's Paradox* (Efron and Morris, 1977): The biasing of an estimate toward an arbitrary low-dimensional target can consistently outperform a less constrained estimate.

Physiological interpretation and future directions

We showed that among several models a sparse network of linear interactions with several latent inputs yielded the best estimates of the noise covariance matrix for cortical microcircuits. This finding is valuable in itself: improved estimates of the noise covari-

ance matrix for large datasets are important in order to understand the role of noise correlations in population coding (Abbott and Dayan, 1999; Sompolinsky et al., 2001; Averbeck et al., 2006; Josic et al., 2009; Ecker et al., 2011)

Moreover, this estimation approach provides a graphical representation of the dependencies in the data that can be used to formulate and test hypotheses about the structure of connectivity in the microcircuit. Importantly, the inferred functional interactions differed substantially from the network of the most significant correlations. For example, the $C_{\text{sparse+latent}}$ estimator reveals a large number of negative interactions that were not present in the sample correlation matrix (Fig. 6 F) and may reflect inhibitory circuitry.

Distances between cells in physical space and in sensory feature space had a stronger effect on the partial correlations estimated by the $C_{\text{sparse+latent}}$ estimator than on sample correlations (Fig. 8 A–C). These differences support the idea that correlations are built up from partial correlations in chains of intermediate cells positioned closer and tuned more similarly to one another, with potentially closer correspondence to anatomical connectivity. These differences may also be at least partially explained by a trivial effect of regularization: the L_1 penalty applied by the estimator (Eq. 18) suppresses small partial correlations to greater extent than large partial correlations, enhancing the apparent effect of distance and tuning. Still, the distinct positive and negative connectivity patterns (Fig. 8 D–F) may reflect geometric and graphical features of local excitatory and inhibitory networks. Indeed, the relationships between patterns of positive and negative connectivities inferred by the estimator resembled the properties of excitatory and inhibitory synaptic connectivities with respect to distance, cortical layers, and feature tuning (Song et al., 2005; Oswald and Reyes, 2008; Adesnik and Scanziani, 2010; Perin et al., 2011; Fino and Yuste, 2011; Hofer et al., 2011; Isaacson and Scanziani, 2011; Levy and Reyes, 2012). For example, while excitatory neurons form synapses within highly

specific local cliques (Perin et al., 2011), inhibitory interneurons form synapses with nearly all excitatory cells within local microcircuits (Fino and Yuste, 2011; Hofer et al., 2011; Packer and Yuste, 2011). To further investigate the link between synaptic connectivity and inferred functional connectivity, in future experiments, we will use molecular markers for various cell types with follow-up multiple whole-cell *in vitro* recordings (Hofer et al., 2011; Ko et al., 2013) to directly compare the inferred functional connectivity graphs to the underlying anatomical circuitry. Finally, the latent units inferred by the estimator can be analyzed for their physiological functions. For example, these latent units may be modulated under different brain states (e.g. slow-wave sleep, attention) and stimulus conditions (e.g. certain types of stimuli may engage feedback connections) (Reimer et al., 2014; Fu et al., 2014).

2.3 Materials and Methods

Ethics statement

All procedures were conducted in accordance with the ethical guidelines of the National Institutes of Health and were approved by the Baylor College of Medicine IACUC.

Surgery and two-photon imaging

The surgical procedures and data acquisition were performed as described in (Cotton et al., 2013): C57BL/6J mice (aged p40–60) were used. For surgery, animals were initially anesthetized with isoflurane (3%). During the experiments, animals were sedated with a mixture of fentanyl (0.05 mg/kg), midazolam (5 mg/kg), and medetomidine (0.5 mg/kg), with boosts of half the initial dose every 3 hours. A craniotomy was performed over the right primary visual cortex. Membrane-permeant calcium indicator Oregon Green 488

BAPTA-1 AM (OGB-1, Invitrogen) was loaded by bolus injection. The craniotomy was sealed using a glass coverslip secured with dental cement.

Calcium imaging began 1 hour after dye injection. All imaging was performed using 3D-RAMP two-photon microscopy (Cotton et al., 2013). First, a 3D stack was acquired and cells were manually segmented. Then calcium signal were collected by sampling in the center of each cell at rates of 100 Hz or higher, depending on the number of cells.

Visual stimulus

The visual stimulus consisted of full-field drifting gratings with 90% contrast, luminance of 10 cd/m², spatial frequency of 0.08 cycles/degree, and temporal frequency of 2 cycles/s. Two types of stimuli were presented for each imaging site: First, directional tuning was mapped using a pseudo-random sequence of drifting gratings at sixteen directions of motion, 500 ms per direction, without blanks, with 12–30 trials for each direction of motion. Second, to measure correlations, the stimulus was modified to include only two directions of motion (in 9 datasets) or five directions (in 22 datasets) and the gratings were presented for 1 second and were separated by 1-second blanks, with 100–300 trials for each direction of motion.

Data processing

All data were processed in MATLAB using the DataJoint data processing chain toolbox (<http://datajoint.github.com>).

The measured fluorescent traces were deconvolved to reconstruct the firing rates for each neuron: First, the first principal component was subtracted from the raw traces in order to reduce common mode noise related to small cardiovascular movements (Cotton et al., 2013). The resulting traces were high-pass filtered above 0.1 Hz and down-

sampled to 20 Hz (Fig. 3 C). Then, the firing rates were estimated using by nonnegative deconvolution (Vogelstein et al., 2010).

Orientation tuning was computed by fitting the mean firing rates for each direction of motion ϕ using two-peaked von Mises tuning functions $f(\phi) = a + b \exp\left[\frac{1}{w}(\cos(\phi - \theta) - 1)\right] + c \exp\left[\frac{1}{w}(\cos(\phi - \theta + \pi) - 1)\right]$ where $b \geq c$ are the amplitudes of the two respective peaks, w is the tuning width, and θ is the preferred direction. The significance of the fit was determined by the permutation test: the labels of the direction were randomly permuted 10,000 times; the p -values of the fits were computed as the fraction of permutations that yielded R^2 equal to or higher than that of the original data. Cells were considered tuned with $p < 0.05$.

For covariance estimation, the analysis was limited to the period with 2 or 5 stimulus conditions and lasted between 14 and 27 minutes (mean 22 minutes). Cells that did not have substantial spiking activity (those whose variance was less than 1% of the median across the site) or whose activity was unstable (those whose variance in the least active quarter of the recording did not exceed 1% of the variance in the most active quarter) were excluded from the analysis.

Cross-validation

To compare the performance of the estimators, we used conventional 10-fold cross-validation: Trials were randomly divided into 10 subsets with approximately equal numbers of trials of each condition in each subset. Each subset was then used as the testing sample with the rest of the data used as the training sample for estimating the covariance matrix. The average validation loss over the 10 folds was reported.

Since each of the regularized estimators had one or two hyperparameters, we used *nested cross-validation*: The outer loop evaluated the performance of the estimators

with the hyperparameter values optimized by cross-validation within the inner loop. Hyperparameters were optimized by a two-phase search algorithm: random search to find a good starting point for the subsequent pattern search to find the global minimum. The inner cross-validation loop subdivided the training dataset from the outer loop to perform 10-fold cross-validation in order to evaluate each choice of the hyperparameter values. Thus the size of the training dataset within the inner loop comprised 81% of the entire recording. Fig. S1 illustrates the dependence of the validation loss on the hyperparameters of the $C_{\text{sparse+latent}}$ estimator for the example site shown in Figures 3 and 6 and the optimal value found by the pattern search algorithm.

When the validation loss was not required, only the inner loop of cross-validation was used on the entire dataset. This approach was used to compute the covariance matrix estimates and their true loss in the simulation study (Fig. 1 Rows 4 and 5) and to analyze the partial correlation structure of the $C_{\text{sparse+latent}}$ estimator (Fig. 6–8).

Covariance estimation

Within the inner loop of cross-validation, regularized covariance matrix estimation required only the sample covariance matrix C_{sample} of the training dataset and the hyperparameter values provided by the outer loop.

Estimator C_{diag} (Eq. 5) used two hyperparameters: the covariance shrinkage intensity $\lambda \in [0, 1]$ and variance shrinkage intensity $\alpha \in [0, 1]$. The variances (the diagonal of C_{sample}) were shrunk linearly toward their mean value $\frac{1}{p} \text{tr}(C_{\text{sample}})$:

$$D = (1 - \alpha)\text{diag}(C_{\text{sample}}) + \alpha \frac{1}{p} \text{tr}(C_{\text{sample}}) I \quad (13)$$

The C_{diag} estimate was then obtained by shrinking C_{sample} toward D according to Eq. 5.

In estimator C_{factor} (Eq. 6), the low-rank matrix L and the diagonal matrix D are

found by solving the minimization problem

$$(L, D) = \underset{\hat{L}, \hat{D}}{\operatorname{arg\,min}} \mathcal{L}(\hat{L} + \hat{D}, C_{\text{sample}}), \quad (14)$$

by an expectation-maximization (EM) algorithm with specified rank of L . In addition, the diagonal matrix of individual variances is shrunk toward its mean value according to Eq. 6.

In estimator C_{sparse} (Eq. 7), the sparse precision matrix S is found by minimizing the L_1 -penalized loss with regularization parameter λ :

$$S = \underset{\hat{S} > 0}{\operatorname{arg\,min}} \mathcal{L}(\hat{S}^{-1}, C_{\text{sample}}) + \lambda \|\hat{S}\|_1 \quad (15)$$

where $\hat{S} > 0$ denotes the constraint that \hat{S} be a positive-definite matrix and $\|\hat{S}\|_1$ is the element-wise L_1 norm of the matrix \hat{S} . This problem formulation is known as *graphical lasso* (Meinshausen and Bühlmann, 2006; Friedman et al., 2008). To solve this minimization problem, we adapted the alternative-direction method of multipliers (ADMM) (Ma et al., 2013). Unlike C_{diag} and C_{factor} , this estimator does not include linear shrinkage: the selection of the sparsity level provides sufficient flexibility to fine-tune the regularization level.

Estimator $C_{\text{sparse+latent}}$ (Eq. 8) estimates a larger sparse precision matrix S^* of the joint distribution of the p observed neurons and d latent units.

$$S^* = \begin{pmatrix} S & S_{12} \\ S_{12}^T & S_{22} \end{pmatrix}, \quad (16)$$

where the $p \times p$ partition S corresponds to the visible units. Then the covariance matrix

of the observed population is

$$C_{\text{sparse+latent}} = \left(S - S_{12} S_{22}^{-1} S_{12}^T \right)^{-1} \quad (17)$$

The rank of the $p \times p$ matrix $L = S_{12} S_{22}^{-1} S_{12}^T$ matches the number of the latent units in the joint distribution. Rather than finding S_{12} and S_{22} separately, L can be estimated as a low-rank positive semidefinite matrix. To simultaneously optimize the sparse component S and the low-rank component L , we adapted the loss function with L_1 penalty on S combined with a penalty on the trace of L (Chandrasekaran et al., 2010; Ma et al., 2013):

$$(S, L) = \underset{\hat{S}, \hat{L}}{\operatorname{argmin}} \mathcal{L}((\hat{S} - \hat{L})^{-1}, C_{\text{sample}}) + \alpha \|\hat{S}\|_1 + \beta \operatorname{tr}(\hat{L}) \quad (18)$$

The trace of a symmetric semidefinite matrix equals the sum of the absolute values of its eigenvalues, *i.e.* its *nuclear norm*; penalty on $\operatorname{tr}(L)$ favors solutions with few non-zero eigenvalues or, equivalently, low-rank solutions while keeping the convexity of the overall optimization problem (Fazel, 2002; Recht et al., 2010). This allows using convex optimization algorithm such as ADMM to be applied with great computational efficiency (Ma et al., 2013).

The partial correlation matrix (Eq. 4) computed from $C_{\text{sparse+latent}}$ includes interactions between the visible and latent units and was used in Fig. 6 C and D and Fig. 7 C, and Fig. 8 A–C). The partial correlation matrix computed from S alone expresses strengths of pairwise interactions

$$P_{\text{sparse}} = -(\operatorname{diag}(S))^{-\frac{1}{2}} S (\operatorname{diag}(S))^{-\frac{1}{2}} \quad (19)$$

and were used in Fig. 6 F, G, H.

The MATLAB code for these computations is available online at <http://github.com/atlab/covest>.

Cross-validation with conditioned variances

Special attention was given to estimating the variances. All evaluations and optimization in this study were defined with respect to the covariance matrices. However, neuroscientists often estimate a common correlation matrix across multiple stimulus conditions when the variances of responses are conditioned on the stimulus (Vogels et al., 1989; Ponce-Alvarez et al., 2013). In this study, we too conditioned the variances on the stimulus but estimated a single correlation matrix across all conditions. Here we describe the computation of the validation loss (Eq. 10) when the variances were allowed to vary with the stimulus condition.

Let T_c and T'_c denote the time bin indices for the training and testing samples, respectively, limited to condition c . Here, the prime symbol marks quantities estimated from the testing sample.

Similar to Eq. 2, the training and testing sample covariance matrices for condition c are

$$C_{c,\text{sample}} = \frac{1}{n_c} \sum_{t \in T_c} (x(t) - \bar{x}_c) (x(t) - \bar{x}_c)^T \quad (20)$$

and

$$C'_{c,\text{sample}} = \frac{1}{n'_c} \sum_{t \in T'_c} (x(t) - \bar{x}_c) (x(t) - \bar{x}_c)^T \quad (21)$$

Here n_c and n'_c denote the sizes of T_c and T'_c , respectively.

Note that $\bar{x}_c = \frac{1}{n_c} \sum_{t \in T_c} x(t)$ is estimated from the training sample but used in both estimates, making $C'_{c,\text{sample}}$ an unbiased estimate of the true covariance matrix, Σ . As such, $C'_{c,\text{sample}}$ can be used for validation.

The common correlation matrix R_{sample} is estimated by averaging the condition-

specific correlations:

$$R_{\text{sample}} = \frac{1}{n} \sum_c n_c \left(V_{c,\text{sample}}^{-\frac{1}{2}} C_{c,\text{sample}} V_{c,\text{sample}}^{-\frac{1}{2}} \right) = \frac{1}{n} \sum_c \sum_{t \in T_c} z(t) z(t)^T, \quad (22)$$

where $n = \sum_c n_c$ and $V_{c,\text{sample}} = \text{diag}(C_{c,\text{sample}})$ is the diagonal matrix containing the sample variances. Then R_{sample} is simply the covariance matrix of the z -score signal $z(t) = V_{c,\text{sample}}^{-\frac{1}{2}} (x(t) - \bar{x}_c)$ of the training sample.

For consistency with prior work, we applied regularization to covariance matrices rather than to correlation matrices. The common covariance matrix was estimated by scaling R_{sample} by the average variances across conditions $V_{\text{sample}} = \frac{1}{n} \sum_c n_c V_{c,\text{sample}}$:

$$C_{\text{sample}} = V_{\text{sample}}^{\frac{1}{2}} R_{\text{sample}} V_{\text{sample}}^{\frac{1}{2}} \quad (23)$$

Note that C_{sample} differs from the sample covariance matrix computed without conditioning the variances on c and this computation helps avoid any biases that would be introduced by ignoring changes in variance.

The covariance matrix estimators C_{diag} , C_{factor} , C_{sparse} or $C_{\text{sparse+latent}}$ convert C_{sample} into its regularized counterpart denoted here as C_{reg} .

To evaluate the estimators, we regularized the conditioned variances by linear shrinkage toward their mean value across all conditions. This was done by scaling C_{reg} by the conditioned variance adjustment matrix $Q_c = \delta I + (1 - \delta) V_{\text{sample}}^{-1} V_{c,\text{sample}}$ to produce the conditioned regularized covariance matrix estimate:

$$C_{c,\text{reg}} = Q_c^{\frac{1}{2}} C_{\text{reg}} Q_c^{\frac{1}{2}} \quad (24)$$

The variance regularization parameter $\delta \in [0, 1]$ was optimized in the inner loop of cross-validation along with the other hyperparameters.

The overall validation loss is obtained by averaging the validation losses across all conditions:

$$\frac{1}{\sum_c n'_c} \sum_c n'_c \mathcal{L}(C_{c,\text{reg}}, C'_{c,\text{sample}}) \quad (25)$$

With negative normal log-likelihood as the validation loss (Eq. 10) and the unbiased validation covariance matrix $C_{c,\text{sample}}$, the loss function in Eq. 25 is an unbiased estimate of the true loss. Hence, it was used for evaluations reported in Fig. 4.

Simulation

For simulation, ground truth covariance matrices were produced by taking 150 independent samples from an artificial population of 50 independent, identically normally distributed units. The covariance matrices were then subjected to the respective regularizations to produce the ground truth matrices for the simulation studies (Fig. 1 Row 2). Samples were then drawn from multivariate normal distributions with the respective true covariance matrices to be estimated by each of the estimators.

Chapter 3

Differential functional connectivity of parvalbumin-positive interneurons

3.1 Introduction

Inhibitory parvalbumin-expressing (PV+) interneurons represent a relatively small fraction of cortical neurons but play critical roles in coordinating the activity of cortical microcircuits. By targeting the perisomatic or axonal domains of nearly all principal cells in their near neighborhood within the same cortical layer and receiving reciprocal input from the same population, PV+ cells are uniquely positioned to exert strong shunting inhibition on all local cells, synchronizing their activity and orchestrating oscillations. The subcellular distribution of their K⁺ and Na⁺ active conductances makes their design optimal for high speed, temporal precision, and selectivity to spatially distributed rather than concentrated input. These properties suggest their role as network stabilizers. They have also been shown to mediate local network computations such as feedback and forward inhibition, divisive input normalization, and “winner-take-all” computations. For a thorough review of the design and function of PV+ interneurons, refer to Hu et al. (2014).

The functional connectivity of PV+ neurons within cortical microcircuits has only recently been studied by Hofer et al. (2011). Using two-photon imaging of calcium signals and genetically encoded fluorescent markers of PV+ cells, they found stronger correlations between the activity of PV+ cells than between PV- cells in local populations of neurons. The patterns of correlations, *i.e.* their relative magnitudes, were also more preserved across stimulus condition between PV+ cells than between PV- cells and between PV+/PV- pairs. These results are consistent with a distinct role of PV+ neurons in stabilizing and orchestrating the population response of the local circuit.

We hypothesized that improved measures of functional connectivity such as the partial correlation estimation developed in Yatsenko et al. (in review) and described in Chapter 2 would provide greater differentiation of the functional connectivity of PV+ cells

than the noise correlations obtained by Hofer et al. (2011) and that these measurements would suggest specific physiological interactions of the PV+ interneuron network. To test this hypothesis, we used the stimulus design and signal acquisition protocol from Yatsenko et al. (in review) in mice with PV+ neurons labeled with the genetically encoded red fluorescent protein tdTomato. We found significantly greater differentiation of the functional connectivity of PV+ cells using regularized partial noise correlations compared to conventional noise correlations. Surprisingly, PV+ neurons exhibited strong positive partial correlations between themselves consistent with either strong and targeted direct interactions or specific patterns of shared input from outside the recorded circuit. Partial correlations between PV– and PV+ cells decreased quickly with physical distance and, for distances greater than 100 μm , became weakly negative, consistent with lateral inhibition of neighboring circuits. Partial correlations had stronger and clearer dependencies than noise correlations on differences in orientation tuning. This was true for both PV-/PV+ and PV-/PV- cell pairs without clear differences between them. Overall, regularized partial correlations produced a clearer picture of functional connectivity than conventional noise correlations, suggesting hypotheses for their physiological origin to be tested in future experiments.

3.2 Results

3.2.1 Networks of partial correlations

Using the same protocol for calcium signal acquisition and stimulus presentations as described in Chapter 2, we estimated the noise correlation structure in the superficial layers of visual cortex in 11 scans from 4 mice with PV+ interneurons genetically labeled with a red fluorescent marker. Each site contained 7–17 interneurons or 2%–10.3% of

the recorded cells. However, these may not be unbiased estimates of the fractions of PV+ neurons because imaged sites were chosen while PV+ cells were visible.

The recorded population sizes in these sites ranged from 137 neurons to 385 (mean 217). As before, the data were binned in 150 ms intervals aligned on stimulus trial onsets; conventional partial correlations matrices were computed and regularized partial correlation matrices were estimated as described in Chapter 2. As before, the partial correlation matrix estimates comprised a sparse pairwise component and a low-rank latent unit component. The sparse component connectivities amounted to 6.9%–26.8% (mean 14.8%) and the number of latent units varied between 39 and 88 (mean 62.3).

Figure 9 depicts the sparse component connectivity for three of the eleven sites.

3.2.2 Functional connectivities by cell pair type

Next, we compared average correlations for both conventional noise correlations and regularized partial correlation matrices conditioned on the cell pair type: non-PV to non-PV (−/−), non-PV to PV (−/+), PV to PV (+/+) (Fig. 10 A and B). Consistent with Hofer et al. (2011), PV+/+ pairs had highest mean noise correlations. The mean +/+ noise correlations were ×1.54 higher than −/− in unpaired comparisons, but with $n = 11$ sites this difference did not reach significance ($p=0.08$, signed-rank test) due to variability between sites. Both +/+ and −/− were significantly higher than −/+: Mean +/+ noise correlations were ×2.9 higher than +/− on average ($p=0.003$) whereas −/− noise correlations were ×1.9 higher than −/+.

Similar to our results in the Chapter 2, regularized partial correlations were lower and most consistent (less dispersed) across sites than sample correlations (Fig. 10 B). More importantly, the cell pair types were better differentiated with partial noise correlations: The average +/+ partial correlations were ×4.9 higher than −/− and ×8.1 higher than −/+.

All comparisons were significant ($p<0.001$, signed rank). In the language of detection theory, noise correlations differentiated $+/+$ site averages from $-/-$ and $-/+$ site averages with d-prime values of 0.88 and 1.7, respectively. The partial noise correlations yielded much higher respective d-prime values of 3.8 and 4.2.

To examine the connectivity rates (connection probabilities) in the sparse network of significant correlations, the noise correlations were thresholded to match the sparsity of the sparse component of the partial correlation matrix estimates (Fig. 10 C and D). Regularized partial correlation estimates provided clearer distinction between the cell pair types yielding consistent relationships between the cell pair types (Fig. 10 D). PV $+/+$ pairs had severalfold higher rates of positive connectivity ($\times 3.1$ higher than $-/+$ and $\times 2.2$ higher than $-/-$). PV $-/-$ pairs also had the lowest rates of negative connectivity ($\times 2.2$ lower than $-/+$ and $\times 2.7$ lower than $-/+$). Notably, $-/-$ pairs had higher rates of positive and negative connectivity than $-/+$ pairs, but the ratio of negative connection rates to positive connection rates was highest for $-/+$ pairs (Fig. 10 F). All these effects were nearly perfectly consistent between the eleven sites when using partial correlation matrix estimates, yielding signed-rank test $p<0.001$ even with this small sample size.

Thresholded noise correlations provided much less consistent results with smaller effect sizes (Fig. 10 C and E).

3.2.3 Distance dependence of functional connectivity by cell pair type

We also compared how functional connectivity depended on the distance between cell soma for each cell pair type (Fig. 11). We found that partial noise correlations (Fig. 11 B) yielded clearer effects with greater consistency across sites than conventional noise correlations (Fig. 11 A). While noise correlations remained high and positive for PV $-/-$ cell pairs, partial noise correlations quickly decayed with distance, suggesting a shorter

Figure 9: Examples of functional connectivity in populations with labeled PV+ interneurons

A. Populations neurons in superficial layers of visual cortex from three mice. The connectivity plot is generated using the same algorithm and color codes as in Fig. 3D and 6G. The edges between nodes indicate the sparse component of the partial correlation matrix. The latent units and correlations induced by them are not shown.

B. The PV+ neurons and the sparse partial correlations between them in the same imaged sites.

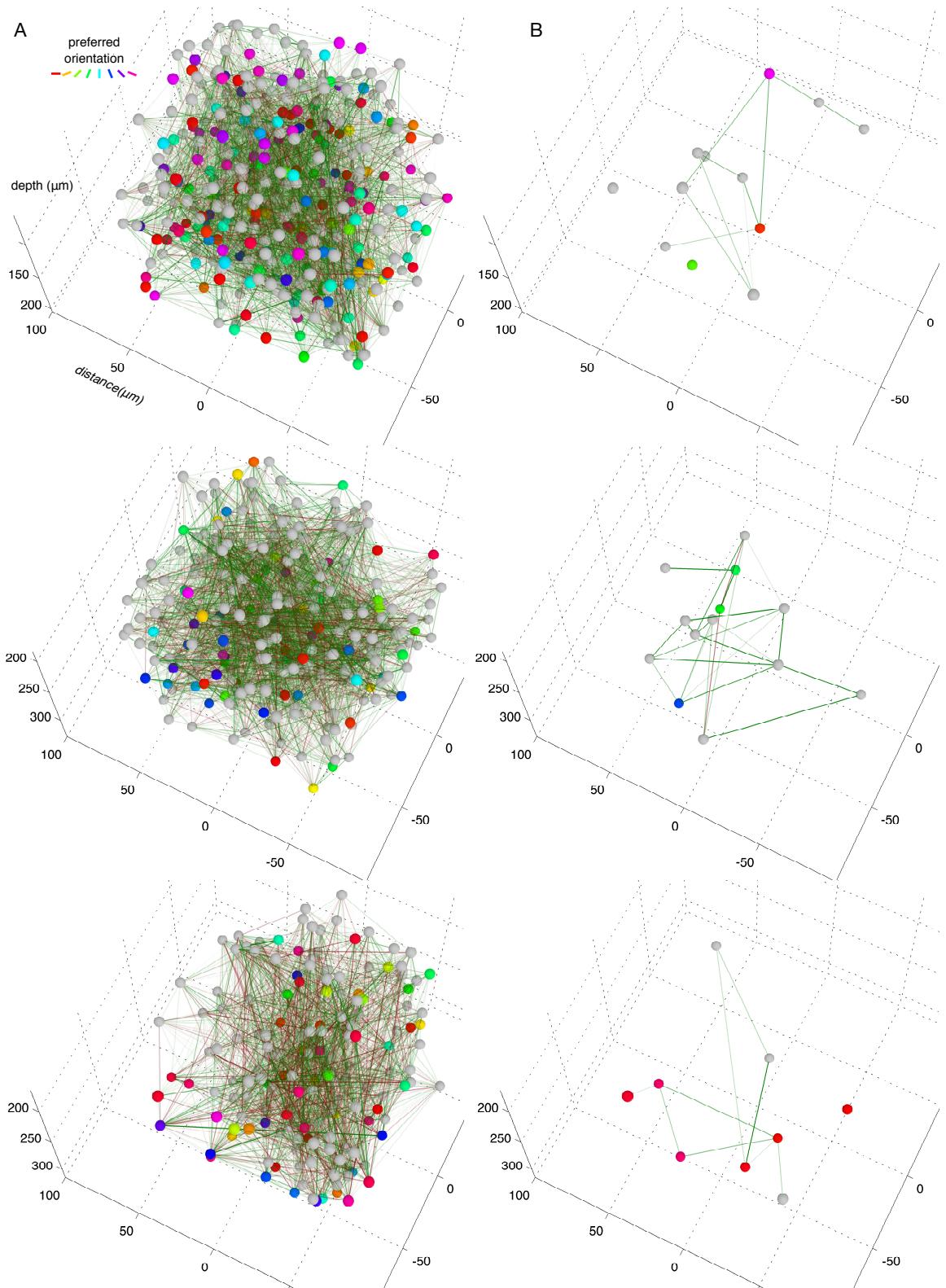
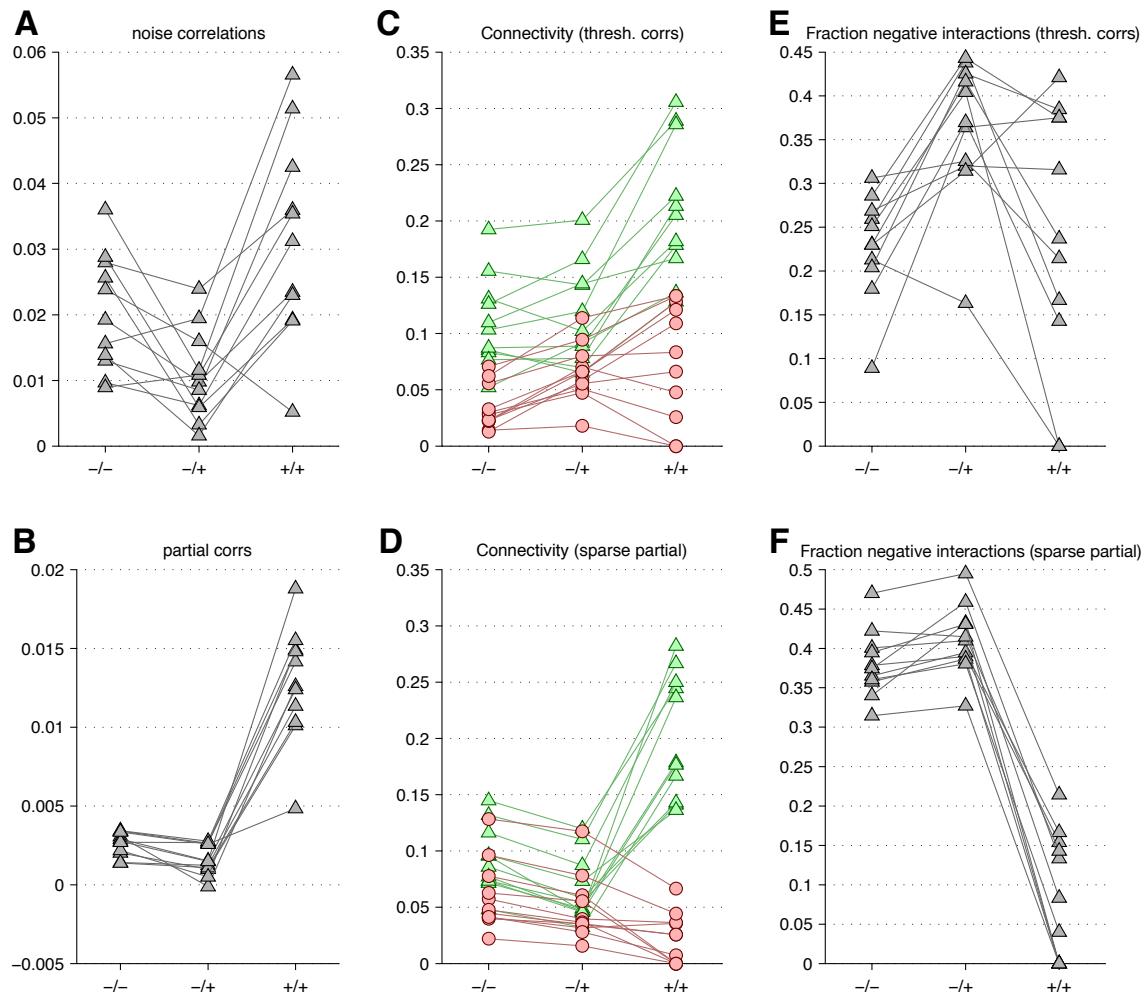


Figure 10: Functional connectivities for PV^{-/-}, PV^{-/+}, and PV^{+/+} cell pairs

Top panels (A,C, and E) show aspects of functional connectivity expressed through conventional noise correlations. Bottom panels (B, D, and F) show connectivity expressed through regularized partial noise correlations. Data points represent averages for each of $n = 11$ sites conditioned on cell pair type. Overall, partial noise correlations provide stronger effects and greater discriminability of cell pair types.

- A.** Average noise correlations.
- B.** Average partial noise correlations with regularization. The relationship to the cell pair types is clearer and more consistent for average partial correlations than for noise correlations.
- C.** Rates of positive (green) and (negative) connectivity obtained by thresholding the correlations to make a sparse matrix of interactions. The sparsities was matched to matrices in panel D.
- D.** Rates of positive (green) and (negative) connectivity in the sparse component of the partial correlation estimates.
- E.** Fraction of negative interactions when interactions are obtained by thresholding noise correlations as in panel C.
- F.** Fraction of negative interactions when interactions are obtained by from the sparse component of the sparse+latent covariance estimator as in panel D. The relationship of connectivity rates to the cell pair types is clearer and more consistent in the sparse component of the sparse+latent estimator than for thresholded correlations.



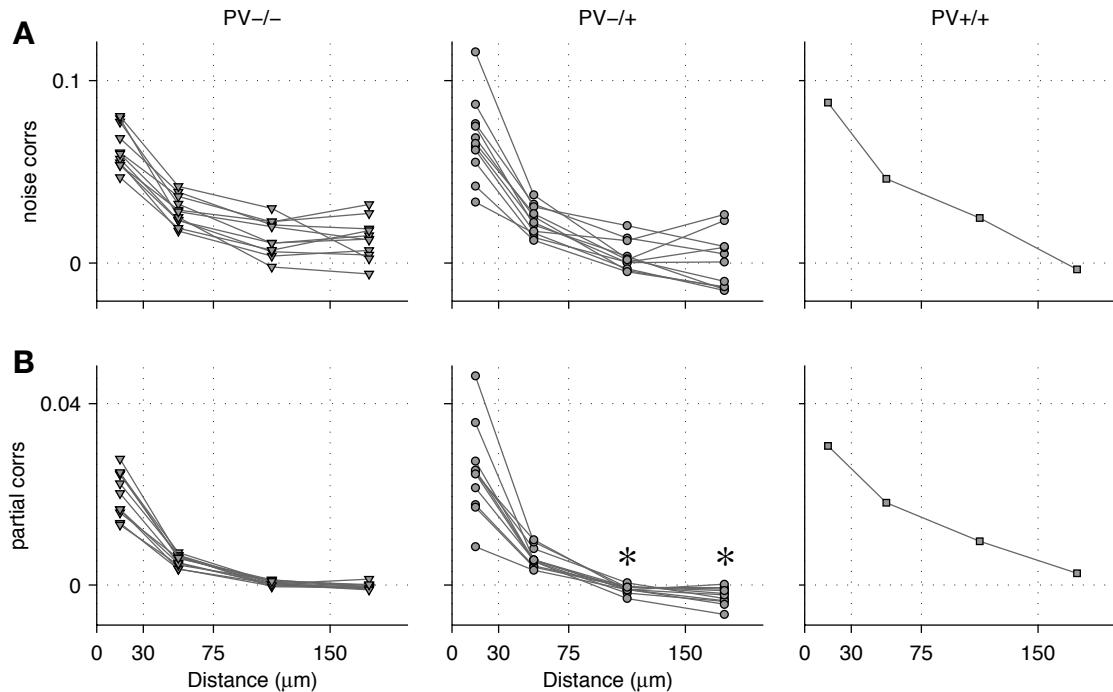


Figure 11: Distance dependence of functional connectivity by cell pair type

A. Average noise correlations binned by distance in 11 sites for PV-/- cell pairs (left), PV-/+ cell pairs (middle), and PV+/+ cell pairs (right, pooled from all sites).

B. Average regularized partial noise correlations binned by distance in 11 sites for PV-/- cell pairs (left), PV-/+ cell pairs (middle), and PV+/+ cell pairs (right, pooled from all sites). The differentiation between cell pair types is stronger for partial correlations than noise correlations. For PV-/+ pairs, the average partial correlations were negative ($p<0.005$, signed rank test) for intersomatic distances greater than 75 μm . In contrast, the partial correlations for PV-/- cell pairs dropped to zero in the same range.

range of more direct interactions. Interestingly, PV-/+ cells pairs had high average partial correlations for short distances but for distances greater than 75 μm , they became reliably negative ($p<0.005$), suggesting lateral inhibition by PV cells of non-PV cells in adjacent circuits.

3.2.4 Orientation tuning dependence of functional connectivity by cell type

We also compared how functional connectivity depended on the difference in preferred orientations (Δori) between cell pairs of each type (Fig. 12). We found that partial noise correlations (Fig. 12 B) yielded clearer effects with greater consistency across sites than conventional noise correlations (Fig. 12 A). Interestingly, partial correlations between PV-- pairs were strongly dependent on Δori ($p<0.04$ for both comparisons of adjacent bins) but not between PV-/PV+ pairs. This may suggest that while excitatory interactions are sensitive to the single-cell response properties, PV+ cells receive excitatory input less selectively.

3.3 Discussion

Although this chapter presents preliminary results with only 11 analyzed datasets, several intriguing findings begin to emerge. First, we confirmed that regularized partial correlations had stronger correspondence to circuit anatomy than conventional noise correlation when compared by cell type (Fig. 10), physical distance between the soma (Fig. 11), and differences in preferred orientations (Fig. 12). Greater correspondence to anatomy constitutes indirect evidence that partial correlations in densely recorded microcircuit provide a closer approximation of specific direct interactions while corre-

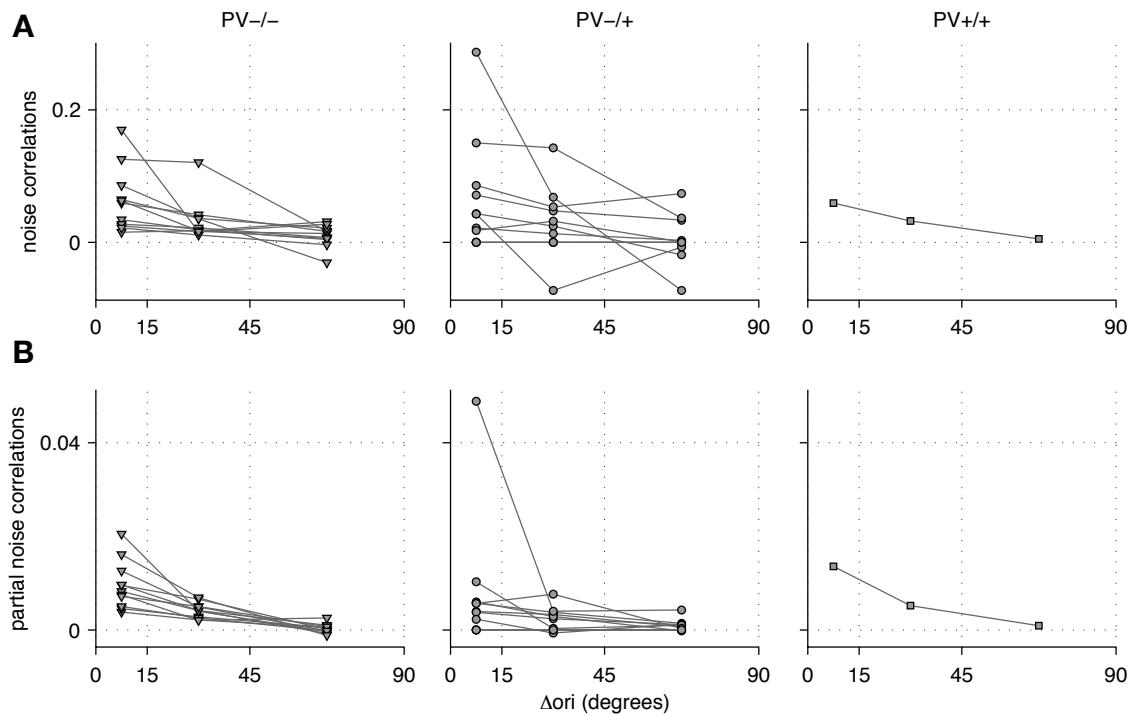


Figure 12: Orientation tuning dependence of functional connectivity by cell pair type
 Each data point represents the average over a site, $n = 9$ sites, except for PV^{+/+} pairs where the data were pooled from all sites.

A. Average noise correlations binned by difference in preferred orientation in 9 sites for PV^{-/-} cell pairs (left), PV^{-/+} cell pairs (middle), and PV^{+/+} cell pairs (right, pooled from all sites).

B. Average regularized partial noise correlations binned by difference in preferred orientation in 9 sites for PV^{-/-} cell pairs (left), PV^{-/+} cell pairs (middle), and PV^{+/+} cell pairs (right, pooled from all sites).

lations reflect the final confounded result of these interaction.

The finding that PV+ cells form dense networks of positive interactions that are not explained by common fluctuations of the entire circuit (Fig. 10 D) runs against a prevailing view that the PV+ subnetwork is highly synchronized, acting as a single unit. Several physiological mechanisms can be proposed to explain such strong sparse pairwise partial correlations. For example, the PV+ subnetwork could be receiving elaborate (not uniform or strongly divergent) input from layers 5. These interaction could also be mediated by gap junctions previously shown to exist between PV+ cells Tamás et al. (2000).

The finding that PV+/PV+ and PV-/PV- pairs both form stronger interactions than PV-/PV+ pairs (Fig. 10 B, D, and F) suggests that PV+ neurons form a relatively isolated subnetwork.

The finding of robustly negative average partial correlations for PV-/+ pairs separated by more than 100 μm (Fig. 11 B), if further confirmed, would constitute, to our knowledge, the first description of a consistently negative correlation between anatomically identified cell groups. The cortical distance at which such negative correlations emerge could serve to delineate the spatial extent of functional modules in the cortex.

The finding that PV-/PV- partial correlations are selective to Δori while PV-/PV+ are not (Fig. 12 B) is consistent with the pattern of excitatory synaptic connectivity measured by Hofer et al. (2011).

3.4 Methods

3.4.1 Data acquisition and processing

The visual stimulus protocol and data acquisition procedures were identical to those described in Chapter 2. PV+ cells were visually identified on the red channel and marked

during the semiautomatic cell segmentation process.

3.4.2 Animals

My fellow member of Andreas Tolias' lab, Cathryn Cadwell, conducted the breeding and genotyping of mice with fluorescent labelling of PV+ interneurons.

The PV-Cre and Ai9 mouse lines were obtained from the Jackson Laboratory (Stock #017320 and #007909, respectively) and maintained as separate homozygous lines in our lab. To generate double heterozygous animals carrying one copy of each allele, homozygous PV-Cre males were crossed with homozygous Ai9 females. Two males and two females aged 66, 67, 68, and 235 days were used in these experiments.

Some animals were genotyped to confirm that they carried the appropriate alleles. Tail samples for genotyping were taken under isoflurane anesthesia on or before p28. DNA extraction was performed in 25 mM NaOH, 0.2 mM EDTA solution at 95°C for 30 minutes followed by neutralization with an equal volume of 40 mM Tris HCl. PCR was carried out using primers for Cre (Cre F: 5'-GCATTACCGGTCGATGCAACGAGTGATGAG-3'; Cre R: 5'-GAGTGAACGAACCTGGTCGAAATCAGTGCG-3') or primers for Ai9 (Ai9 F: 5'-GTAATGCAGAAGAAGACTATGGGCTGGGAG-3'; Ai9 R: 5'-ATGTCCAGCTTGGAGTCC-ACGTAGTAGTAG-3') using a thermal cycler (Eppendorf Mastercycler pro S) with the following program: 94°C for 50 s; 30 cycles of 94°C for 20 s, 60°C for 45 s, and 68°C for 2 min; 68°C for 1 min; and 10°C ∞.

Chapter 4

Conclusions and future directions

The general aim of this dissertation was to understand the principles of organization of the direct, mechanistic interactions between neurons in cortical circuits and to understand how their contribution to the overall function of the neural circuit such as sensation and movement. Since interactions between neurons are mediated by synaptic connections, one broad approach to this problem is the study of connectomics using anatomical techniques (Denk and Horstmann, 2004; Perin et al., 2011; Jiang et al., 2013). Another broad approach is to characterize interactions in neuronal populations through statistical descriptions of the observed activity in the circuit *in vivo* (Perkel et al., 1967; Zohary et al., 1994; Bair et al., 2001; Smith and Kohn, 2008; Ecker et al., 2010). This dissertation aimed to advance such statistical descriptions and to show their relevance by demonstrating a strong relationship to the anatomical properties of circuit organization.

When statistical descriptions of populations activity are presented, inevitably, the question arises, “But what does this mean?” Do the interactions in functional connectivity models translate to anatomical, synaptic connections and if not, then what? Sometimes an improved statistical model of population activity simply means that we can more reliably read out the information encoded by the population (Pillow et al., 2011). However, in this dissertation, we specifically set out to measure statistical expressions of interactions between cells.

Instructive in this regard is the success story of receptive field and feature selectivity of single cells. A fortuitous fact of neurobiology is that the spiking activity of individual cells can be efficiently described as a function of external stimuli by simple statistical models with only a handful of parameters (Carandini et al., 2005). From the original discovery and with subsequent refinement of such descriptions over the past 50+ years, constructs such as *receptive fields* and *orientation tuning* have become principal tools for defining the *functional architecture* of sensory cortex, *i.e.* the relationship between single-cell properties related to neural function and the cell types, cortical layers, synap-

tic connectivity, and the physical arrangements of cells (Hubel and Wiesel, 1962; Ohki et al., 2005; Reid, 2012). Although receptive fields did suggest a mechanistic explanation, the immediate validation of receptive fields as a fundamental principle of neural computation came from finding intricate relationships between the tuning properties of cells and circuit anatomy. Tuning properties of cells were organized into maps, columns, and pinwheels.

Similarly, a successful metric of functional connectivity need not estimate the exact synaptic connectivity in the circuit from its spiking activity — quite likely an intractable task — but extract a simple succinct summary of the activity that relates function to anatomy, constrains theoretical models of underlying computations, and feeds hypotheses of physiological and computation principles.

Starting with the intuition that conditioning the activity of pairs of variables on all the other measured variables should improve estimation of their conditional independence or absence of direct interaction, we tested the hypothesis that partial correlations, estimated with a numerically stable method, should yield a closer reflection of physiological interactions than the currently prevalent noise correlations. In the first study described in Chapter 2, we performed empirical cross-validation of several regularization schemes to select the optimal estimator of partial correlations for calcium signals in cortical populations.

In the second study (Chapter 3), we found new, robust, and nontrivial relationships between the distributions of partial correlations and the anatomical organization of the local circuit. Some of these effects may have intuitive explanations such as known differential rates of directed inhibitory or excitatory projections. Others may suggest multiple explanations to be resolved in future experiments. In ongoing experiments, we are estimating functional connectivity in cortical populations with labeled somatostatin-positive neurons and ontogenetically related groups of neurons. Relating studies of

functional connectivity to studies of synaptic connectivity and functional architecture of the neocortex will help form a more complete understanding of cortical computations.

Bibliography

- Abbott L and Dayan P (1999) The effect of correlated variability on the accuracy of a population code, *Neural computation* **11**(1), 91–101.
- Adesnik H and Scanziani M (2010) Lateral competition for cortical space by layer-specific horizontal circuits, *Nature* **464**(7292), 1155–1160.
- Aertsen A, Gerstein G, Habib M, and Palm G (1989) Dynamics of neuronal firing correlation: modulation of “effective connectivity”, *Journal of neurophysiology* **61**(5), 900–917.
- Ahrens MB, Orger MB, Robson DN, Li JM, and Keller PJ (2013) Whole-brain functional imaging at cellular resolution using light-sheet microscopy, *Nature methods* **10**(5), 413–420.
- Alivisatos AP, Chun M, Church GM, Deisseroth K, Donoghue JP, Greenspan RJ, McEuen PL, Roukes M, Sejnowski TJ, Weiss PS, et al. (2013) The brain activity map, *Science* **339**(6125), 1284–1285.
- Alonso JM and Martinez LM (1998) Functional connectivity between simple cells and complex cells in cat striate cortex, *Nature neuroscience* **1**(5), 395–403.
- Anderson T (2003) An introduction to multivariate statistical analysis, *Wiley series in probability and statistics*.

Arieli A, Shoham D, Hildesheim R, and Grinvald A (1995) Coherent spatiotemporal patterns of ongoing activity revealed by real-time optical imaging coupled with single-unit recording in the cat visual cortex, *Journal of Neurophysiology* **73**(5), 2072–2093.

Arlot S and Celisse A (2010) A survey of cross-validation procedures for model selection, *Statistics Surveys* **4**, 40–79.

Averbeck BB, Latham PE, and Pouget A (2006) Neural correlations, population coding and computation., *Nat Rev Neurosci* **7**(5), 358–366.

Baba K, Shibata R, and Sibuya M (2004) Partial correlation and conditional correlation as measures of conditional independence, *Australian & New Zealand Journal of Statistics* **46**(4), 657–664.

Bair W, Zohary E, and Newsome WT (2001) Correlated firing in macaque visual area mt: time scales and relationship to behavior, *The journal of Neuroscience* **21**(5), 1676–1697.

Berens P, Ecker AS, Cotton RJ, Ma WJ, Bethge M, and Tolias AS (2012) A fast and simple population code for orientation in primate v1, *The Journal of Neuroscience* **32**(31), 10618–10626.

Berens P, Ecker AS, Gerwinn S, Tolias AS, and Bethge M (2011) Reassessing optimal neural population codes with neurometric functions, *Proc Natl Acad Sci U S A* **108**(11), 4423–8.

Bickel PJ, Li B, Tsybakov AB, van de Geer SA, Yu B, Valdés T, Rivero C, Fan J, and van der Vaart A (2006) Regularization in statistics, *Test* **15**(2), 271–344.

Carandini M, Demb JB, Mante V, Tolhurst DJ, Dan Y, Olshausen BA, Gallant JL, and Rust

NC (2005) Do we know what the early visual system does?, *The Journal of neuroscience* **25**(46), 10577–10597.

Chandrasekaran V, Parrilo PA, and Willsky AS (2010) Latent variable graphical model selection via convex optimization, in *Communication, Control, and Computing (Allerton), 2010 48th Annual Allerton Conference on*, pp. 1610–1613, IEEE.

Chapin JK and Nicolelis MA (1999) Principal component analysis of neuronal ensemble activity reveals multidimensional somatosensory representations, *Journal of neuroscience methods* **94**(1), 121–140.

Chiu C and Weliky M (2002) Relationship of correlated spontaneous activity to functional ocular dominance columns in the developing visual cortex, *Neuron* **35**(6), 1123–1134.

Ch'Ng YH and Reid RC (2010) Cellular imaging of visual cortex reveals the spatial and functional organization of spontaneous activity, *Frontiers in integrative neuroscience* **4**.

Cohen MR and Maunsell JH (2009) Attention improves performance primarily by reducing interneuronal correlations, *Nature neuroscience* **12**(12), 1594–1600.

Cohen MR and Newsome WT (2008) Context-dependent changes in functional circuitry in visual area mt, *Neuron* **60**(1), 162–173.

Cotton RJ, Froudarakis E, Storer P, Saggau P, and Tolias AS (2013) Three-dimensional mapping of microcircuit correlation structure, *Frontiers in Neural Circuits* **7**, 151.

Dempster A (1972) Covariance selection, *Biometrics* pp. 157–175.

Denk W and Horstmann H (2004) Serial block-face scanning electron microscopy to reconstruct three-dimensional tissue nanostructure, *PLoS biology* **2**(11), e329.

Denman DJ and Contreras D (2013) The structure of pairwise correlation in mouse primary visual cortex reveals functional organization in the absence of an orientation map, *Cereb Cortex*.

Douglas RJ and Martin KA (2004) Neuronal circuits of the neocortex, *Annu. Rev. Neurosci.* **27**, 419–451.

Douglas RJ and Martin KAC (2007) Mapping the matrix: the ways of neocortex., *Neuron* **56**(2), 226–238.

Ecker AS, Berens P, Cotton RJ, Subramaniyan M, Denfield GH, Cadwell CR, Smirnakis SM, Bethge M, and Tolias AS (2014) State dependence of noise correlations in macaque primary visual cortex, *Neuron* **82**(1), 235–248.

Ecker AS, Berens P, Keliris GA, Bethge M, Logothetis NK, and Tolias AS (2010) Decorrelated neuronal firing in cortical microcircuits, *Science* **327**(5965), 584–587.

Ecker AS, Berens P, Tolias AS, and Bethge M (2011) The effect of noise correlations in populations of diversely tuned neurons, *The Journal of Neuroscience* **31**(40), 14272–14283.

Efron B and Morris CN (1977) *Stein's paradox in statistics*, WH Freeman.

Fazel M (2002) *Matrix rank minimization with applications*, Ph.D. thesis, Stanford University.

Feldt S, Bonifazi P, and Cossart R (2011) Dissecting functional connectivity of neuronal microcircuits: experimental and theoretical insights, *Trends in neurosciences* **34**(5), 225–236.

Fino E and Yuste R (2011) Dense inhibitory connectivity in neocortex, *Neuron* **69**(6), 1188–203.

Friedman J, Hastie T, and Tibshirani R (2008) Sparse inverse covariance estimation with the graphical lasso, *Biostatistics* **9**(3), 432–441.

Friedman JH (1989) Regularized discriminant analysis, *Journal of the American statistical association* **84**(405), 165–175.

Friston KJ (2011) Functional and effective connectivity: a review, *Brain connectivity* **1**(1), 13–36.

Fu Y, Tucciarone JM, Espinosa JS, Sheng N, Darcy DP, Nicoll RA, Huang ZJ, and Stryker MP (2014) A cortical circuit for gain control by behavioral state, *Cell* **156**(6), 1139–1152.

Ganmor E, Segev R, and Schneidman E (2011) Sparse low-order interaction network underlies a highly correlated and learnable neural population code, *Proc Natl Acad Sci U S A* **108**(23), 9679–84.

Gerstein G and Clark W (1964) Simultaneous studies of firing patterns in several neurons, *Science* **143**(3612), 1325–1327.

Gerstein GL, Bedenbaugh P, and Aertsen AM (1989) Neuronal assemblies, *Biomedical Engineering, IEEE Transactions on* **36**(1), 4–14.

Gerstein GL and Perkel DH (1969) Simultaneously recorded trains of action potentials: analysis and functional interpretation, *Science* **164**(3881), 828–830.

Goard M and Dan Y (2009) Basal forebrain activation enhances cortical coding of natural scenes, *Nat Neurosci* **12**(11), 1444–9.

Gobel W and Helmchen F (2007) In vivo calcium imaging of neural network function., *Physiology (Bethesda)* **22**(December), 358–365.

Golshani P, Gonçalves JT, Khoshkhoo S, Mostany R, Smirnakis S, and Portera-Cailliau C (2009) Internally mediated developmental desynchronization of neocortical network activity, *J Neurosci* **29**(35), 10890–9.

Greenberg DS, Houweling AR, and Kerr JN (2008) Population imaging of ongoing neuronal activity in the visual cortex of awake rats, *Nature neuroscience* **11**(7), 749–751.

Grewe BF, Langer D, Kasper H, Kampa BM, and Helmchen F (2010) High-speed in vivo calcium imaging reveals neuronal network activity with near-millisecond precision, *Nat Meth* **7**(5), 399–405.

Gu Y, Liu S, Fetsch CR, Yang Y, Fok S, Sunkara A, DeAngelis GC, and Angelaki DE (2011) Perceptual learning reduces interneuronal correlations in macaque visual cortex, *Neuron* **71**(4), 750–761.

Hansen BJ, Chelaru MI, and Dragoi V (2012) Correlated variability in laminar cortical circuits, *Neuron* **76**(3), 590–602.

Harris K, Bartho P, Chadderton P, Curto C, de la Rocha J, Hollender L, Itsikov V, Luczak A, Marguet S, Renart A, et al. (2011) How do neurons work together? lessons from auditory cortex, *Hearing research* **271**(1), 37–53.

Harris KD (2005) Neural signatures of cell assembly organization, *Nature Reviews Neuroscience* **6**(5), 399–407.

Harris KD and Mrsic-Flogel TD (2013) Cortical connectivity and sensory coding, *Nature* **503**(7474), 51–58.

Hertz J, Roudi Y, and Tyrcha J (2011) Ising models for inferring network structure from spike data, *arXiv preprint arXiv:1106.1752*.

Hofer SB, Ko H, Pichler B, Vogelstein J, Ros H, Zeng H, Lesica NA, Mrsic-Flogel TD, et al. (2011) Differential connectivity and response dynamics of excitatory and inhibitory neurons in visual cortex, *Nature neuroscience* **14**(8), 1045–1052.

Hu H, Gan J, and Jonas P (2014) Fast-spiking, parvalbumin+ gabaergic interneurons: From cellular design to microcircuit function, *Science* **345**(6196), 1255263.

Hubel DH and Wiesel TN (1962) Receptive fields, binocular interaction and functional architecture in the cat's visual cortex, *J Physiol* **160**, 106–54.

Isaacson JS and Scanziani M (2011) How inhibition shapes cortical activity, *Neuron* **72**(2), 231–243.

Jaynes ET (1957) Information theory and statistical mechanics., *Phys. Rev.* **106**(4), 620–630.

Jiang X, Wang G, Lee AJ, Stornetta RL, and Zhu JJ (2013) The organization of two new cortical interneuronal circuits, *Nature neuroscience* **16**(2), 210–218.

Josic K, Shea-Brown E, Doiron B, and de la Rocha J (2009) Stimulus-dependent correlations and population codes, *Neural computation* **21**(10), 2774–2804.

Katona G, Szalay G, Maák P, Kaszás A, Veress M, Hillier D, Chiovini B, Vizi ES, Roska B, and Rózsa B (2012) Fast two-photon in vivo imaging with three-dimensional random-access scanning in large tissue volumes, *Nat Methods*.

Kenet T, Bibitchkov D, Tsodyks M, Grinvald A, and Arieli A (2003) Spontaneously emerging cortical representations of visual attributes, *Nature* **425**(6961), 954–956.

Ko H, Cossell L, Baragli C, Antolik J, Clopath C, Hofer SB, and Mrsic-Flogel TD (2013) The emergence of functional microcircuits in visual cortex, *Nature* **496**(7443), 96–100.

Ko H, Hofer SB, Pichler B, Buchanan KA, Sjöström PJ, and Mrsic-Flogel TD (2011) Functional specificity of local synaptic connections in neocortical networks, *Nature* **473**(7345), 87–91.

Kohn A and Smith MA (2005) Stimulus dependence of neuronal correlation in primary visual cortex of the macaque, *J Neurosci* **25**(14), 3661–73.

Kohn A, Zandvakili A, and Smith MA (2009) Correlations and brain states: from electrophysiology to functional imaging, *Curr Opin Neurobiol* **19**(4), 434–8.

Köster U, Sohl-Dickstein J, Gray CM, and Olshausen BA (2013) Higher order correlations within cortical layers dominate functional connectivity in microcolumns, *arXiv preprint arXiv:1301.0050*.

Ledoit O and Wolf M (2003) Improved estimation of the covariance matrix of stock returns with an application to portfolio selection, *Journal of Empirical Finance* **10**(5), 603–621.

Ledoit O and Wolf M (2004) A well-conditioned estimator for large-dimensional covariance matrices, *Journal of multivariate analysis* **88**(2), 365–411.

Leung LC, Wang GX, and Mourrain P (2013) Imaging zebrafish neural circuitry from whole brain to synapse, *Frontiers in neural circuits* **7**.

Levy RB and Reyes AD (2012) Spatial profile of excitatory and inhibitory synaptic connectivity in mouse primary auditory cortex, *J Neurosci* **32**(16), 5609–19.

Lopes-dos Santos V, Conde-Ocacionez S, Nicolelis MA, Ribeiro ST, and Tort AB (2011) Neuronal assembly detection and cell membership specification by principal component analysis, *PloS one* **6**(6), e20996.

Lopes-dos Santos V, Ribeiro S, and Tort AB (2013) Detecting cell assemblies in large neuronal populations, *Journal of Neuroscience Methods* **220**(2), 149 – 166, URL <http://www.sciencedirect.com/science/article/pii/S0165027013001489>.

Luczak A, Barthó P, and Harris KD (2009) Spontaneous events outline the realm of possible sensory responses in neocortical populations, *Neuron* **62**(3), 413–25.

Ma S, Xue L, and Zou H (2013) Alternating direction methods for latent variable gaussian graphical model selection, *Neural computation* pp. 1–27.

MacLean JN, Watson BO, Aaron GB, and Yuste R (2005) Internal dynamics determine the cortical response to thalamic stimulation, *Neuron* **48**(5), 811–23.

Malmersjö S, Rebellato P, Smedler E, Planert H, Kanatani S, Liste I, Nanou E, Sunner H, Abdelhady S, Zhang S, et al. (2013) Neural progenitors organize in small-world networks to promote cell proliferation, *Proceedings of the National Academy of Sciences* **110**(16), E1524–E1532.

Meinshausen N and Bühlmann P (2006) High-dimensional graphs and variable selection with the lasso, *The Annals of Statistics* **34**(3), 1436–1462.

Miller JeK, Ayzenshtat I, Carrillo-Reid L, and Yuste R (2014) Visual stimuli recruit intrinsically generated cortical ensembles, *Proceedings of the National Academy of Sciences* p. 201406077.

Moore GP, Segundo JP, Perkel DH, and Levitan H (1970) Statistical signs of synaptic interaction in neurons, *Biophysical Journal* **10**(9), 876–900.

Mountcastle VB (1997) The columnar organization of the neocortex., *Brain* **120** (Pt 4), 701–722.

Nirenberg S and Latham PE (2003) Decoding neuronal spike trains: How important are correlations?, *Proceedings of the National Academy of Sciences* **100**(12), 7348–7353.

Ohki K, Chung S, Ch'ng YH, Kara P, and Reid RC (2005) Functional imaging with cellular resolution reveals precise micro-architecture in visual cortex., *Nature* **433**(7026), 597–603.

Okun M, Yger P, Marguet SL, Gerard-Mercier F, Benucci A, Katzner S, Busse L, Carandini M, and Harris KD (2012) Population rate dynamics and multineuron firing patterns in sensory cortex, *J Neurosci* **32**(48), 17108–19.

Ostojic S, Brunel N, and Hakim V (2009) How connectivity, background activity, and synaptic properties shape the cross-correlation between spike trains, *The Journal of Neuroscience* **29**(33), 10234–10253.

Oswald AMM and Reyes AD (2008) Maturation of intrinsic and synaptic properties of layer 2/3 pyramidal neurons in mouse auditory cortex, *Journal of neurophysiology* **99**(6), 2998.

Packer AM and Yuste R (2011) Dense, unspecific connectivity of neocortical parvalbumin-positive interneurons: a canonical microcircuit for inhibition?, *The Journal of Neuroscience* **31**(37), 13260–13271.

Peng J, Wang P, Zhou N, and Zhu J (2009) Partial correlation estimation by joint sparse regression models, *Journal of the American Statistical Association* **104**(486).

Perin R, Berger TK, and Markram H (2011) A synaptic organizing principle for cortical neuronal groups, *Proc Natl Acad Sci U S A* **108**(13), 5419–24.

Perkel DH, Gerstein GL, and Moore GP (1967) Neuronal spike trains and stochastic point processes: II. simultaneous spike trains, *Biophysical journal* **7**(4), 419–440.

Peyrache A, Benchenane K, Khamassi M, Wiener SI, and Battaglia FP (2010) Principal component analysis of ensemble recordings reveals cell assemblies at high temporal resolution, *Journal of computational neuroscience* **29**(1-2), 309–325.

Pfau D, Pnevmatikakis EA, and Paninski L (2013) Robust learning of low-dimensional dynamics from large neural ensembles, in *Advances in Neural Information Processing Systems*, pp. 2391–2399.

Pillow JW, Ahmadian Y, and Paninski L (2011) Model-based decoding, information estimation, and change-point detection techniques for multineuron spike trains, *Neural Computation* **23**(1), 1–45.

Pillow JW, Shlens J, Paninski L, Sher A, Litke AM, Chichilnisky E, and Simoncelli EP (2008) Spatio-temporal correlations and visual signalling in a complete neuronal population, *Nature* **454**(7207), 995–999.

Ponce-Alvarez A, Thiele A, Albright TD, Stoner GR, and Deco G (2013) Stimulus-dependent variability and noise correlations in cortical mt neurons, *Proceedings of the National Academy of Sciences* **110**(32), 13162–13167.

Recht B, Fazel M, and Parrilo PA (2010) Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization, *SIAM review* **52**(3), 471–501.

Reddy GD and Saggau P (2005) Fast three-dimensional laser scanning scheme using acousto-optic deflectors., *J Biomed Opt* **10**(6), 064038.

Reid RC (2012) From functional architecture to functional connectomics, *Neuron* **75**(2), 209–217.

Reimer J, Froudarakis E, Cadwell CR, Yatsenko D, Denfield GH, and Tolias AS (2014) Pupil

fluctuations track fast switching of cortical states during quiet wakefulness, *Neuron* **84**(2), 355–362.

Renart A, de la Rocha J, Bartho P, Hollender L, Parga N, Reyes A, and Harris KD (2010) The asynchronous state in cortical circuits, *Science* **327**(5965), 587–90.

Rosenbaum R and Josić K (2011) Mechanisms that modulate the transfer of spiking correlations, *Neural computation* **23**(5), 1261–1305.

Rothschild G, Nelken I, and Mizrahi A (2010) Functional organization and population dynamics in the mouse primary auditory cortex, *Nature neuroscience* **13**(3), 353–360.

Roudi Y, Nirenberg S, and Latham PE (2009) Pairwise maximum entropy models for studying large biological systems: when they can work and when they can't, *PLoS computational biology* **5**(5), e1000380.

Ryali S, Chen T, Supekar K, and Menon V (2012) Estimation of functional connectivity in fmri data using stability selection-based sparse partial correlation with elastic net penalty, *Neuroimage* **59**(4), 3852–3861.

Sadovsky AJ and MacLean JN (2014) Mouse visual neocortex supports multiple stereotyped patterns of microcircuit activity, *The Journal of Neuroscience* **34**(23), 7769–7777, doi: 10.1523/JNEUROSCI.0169-14.2014, URL <http://www.jneurosci.org/content/34/23/7769.abstract>.

Salinas E and Sejnowski TJ (2001) Correlated neuronal activity and the flow of neural information, *Nature Reviews Neuroscience* **2**(8), 539–550.

Schäfer J, Strimmer K, et al. (2005) A shrinkage approach to large-scale covariance matrix estimation and implications for functional genomics, *Statistical applications in genetics and molecular biology* **4**(1), 32.

Schneidman E, Berry MJ 2nd, Segev R, and Bialek W (2006) Weak pairwise correlations imply strongly correlated network states in a neural population, *Nature* **440**(7087), 1007–12.

Shadlen MN and Newsome WT (1998) The variable discharge of cortical neurons: implications for connectivity, computation, and information coding, *J Neurosci* **18**(10), 3870–96.

Shlens J, Field GD, Gauthier JL, Greschner M, Sher A, Litke AM, and Chichilnisky EJ (2009) The structure of large-scale synchronized firing in primate retina, *J Neurosci* **29**(15), 5022–31.

Smith MA, Jia X, Zandvakili A, and Kohn A (2013) Laminar dependence of neuronal correlations in visual cortex, *Journal of neurophysiology* **109**(4), 940–947.

Smith MA and Kohn A (2008) Spatial and temporal scales of neuronal correlation in primary visual cortex., *J Neurosci* **28**(48), 12591–12603.

Smith MA and Sommer MA (2013) Spatial and temporal scales of neuronal correlation in visual area v4, *The Journal of Neuroscience* **33**(12), 5422–5432.

Sompolinsky H, Yoon H, Kang K, and Shamir M (2001) Population coding in neuronal systems with correlated noise, *Physical Review E* **64**(5), 051904.

Song S, Sjöström PJ, Reigl M, Nelson S, and Chklovskii DB (2005) Highly nonrandom features of synaptic connectivity in local cortical circuits, *PLoS biology* **3**(3), e68.

Stevenson IH, Rebesco JM, Miller LE, and Körding KP (2008) Inferring functional connections between neurons, *Current opinion in neurobiology* **18**(6), 582–588.

Stosiek C, Garaschuk O, Holthoff K, and Konnerth A (2003) In vivo two-photon calcium imaging of neuronal networks, *Proceedings of the National Academy of Sciences* **100**(12), 7319–7324.

Tamás G, Buhl EH, Lörincz A, and Somogyi P (2000) Proximally targeted gabaergic synapses and gap junctions synchronize cortical interneurons, *Nature neuroscience* **3**(4), 366–371.

Tang A, Jackson D, Hobbs J, Chen W, Smith JL, Patel H, Prieto A, Petrusca D, Grivich MI, Sher A, Hottowy P, Dabrowski W, Litke AM, and Beggs JM (2008) A maximum entropy model applied to spatial and temporal correlations from cortical networks in vitro, *J Neurosci* **28**(2), 505–18.

Theis L, Berens P, Froudarakis E, Reimer J, Roman-Roson M, Baden T, Euler T, Tolias AS, and Bethge M (2014) Supervised learning sets benchmark for robust spike detection from calcium imaging signals, *bioRxiv* p. 010777.

Tkačik G, Marre O, Amodei D, Schneidman E, Bialek W, Berry I, and Michael J (2013) Searching for collective behavior in a network of real neurons, *arXiv preprint arXiv:1306.3061*.

Tkacik G, Schneidman E, Berry I, Michael J, and Bialek W (2006) Ising models for networks of real neurons, *arXiv preprint q-bio/0611072*.

Varoquaux G, Gramfort A, Poline JB, and Thirion B (2012) Markov models for fmri correlation structure: is brain functional connectivity small world, or decomposable into networks?, *Journal of Physiology-Paris* **106**(5), 212–221.

Vélez-Fort M, Rousseau CV, Niedworok CJ, Wickersham IR, Rancz EA, Brown AP, Strom

M, and Margrie TW (2014) The stimulus selectivity and connectivity of layer six principal cells reveals cortical microcircuits underlying visual processing, *Neuron*.

Vogels R, Spileers W, and Orban G (1989) The response variability of striate cortical neurons in the behaving monkey, *Experimental brain research* **77**(2), 432–436.

Vogelstein JT, Packer AM, Machado TA, Sippy T, Babadi B, Yuste R, and Paninski L (2010) Fast nonnegative deconvolution for spike train inference from population calcium imaging, *Journal of neurophysiology* **104**(6), 3691–3704.

Whittaker J (1990) *Graphical models in applied multivariate statistics*, Wiley Publishing.

Yatsenko D, Josić K, Ecker AS, Froudarakis E, Cotton RJ, and Tolias AS (in review) Improved estimation and interpretation of correlations in neural circuits, *PLoS Computational Biology*.

Yu S, Huang D, Singer W, and Nikolic D (2008) A small world of neuronal synchrony, *Cereb Cortex* **18**(12), 2891–901.

Zalesky A, Fornito A, and Bullmore E (2012) On the use of correlation as a measure of network connectivity, *Neuroimage* **60**(4), 2096–2106.

Zohary E, Shadlen MN, and Newsome WT (1994) Correlated neuronal discharge rate and its implications for psychophysical performance., *Nature* **370**(6485), 140–143.