

Partial neural correlations suggest detailed interactions in cortical circuits  
 Dimitri Yatsenko,<sup>1,\*</sup>, Krešimir Josić<sup>2</sup>, Alexander S. Ecker<sup>1,3,4</sup>, Emmanouil Froudarakis<sup>1</sup>, R. James Cotton<sup>1</sup>, Andreas S. Tolias<sup>1,5</sup>

**1 Department of Neuroscience, Baylor College of Medicine, Houston, TX, USA**

**2 Department of Mathematics and Department of Biology and Biochemistry, University of Houston, Houston, TX, USA**

**3 Werner Reichardt Center for Integrative Neuroscience and Institute for Theoretical Physics, University of Tübingen, Germany**

**4 Bernstein Center for Computational Neuroscience, Tübingen, Germany**

**5 Department of Computational and Applied Mathematics, Rice University, Houston, TX, USA**

\* E-mail: [yatsenko@cns.bcm.edu](mailto:yatsenko@cns.bcm.edu)

## Abstract

Ambitious projects aim to record the activity of ever larger and denser subsets of neurons *in vivo*. Correlations in neural activity measured in such recordings can reveal important aspects of the organization of neural circuits. However, estimating and interpreting large correlation matrices is 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 of about 200  $\mu\text{m}$  wide and 100  $\mu\text{m}$  deep (150–350 cells) in mouse visual cortex. We hypothesized that in these dense recordings, the correlation matrix should be best represented as the combination of a sparse matrix of pairwise partial correlations representing local interactions and a low-rank component representing common fluctuations and external inputs. Indeed, in cross-validation tests, the covariance matrix estimator with such a structure consistently outperformed other regularized estimators. The connectivity patterns inferred from the sparse component of the estimate were strongly related to the geometrical arrangement and orientation tuning properties of cells: the density of positive ‘excitatory’ connections decreased rapidly with geometric distances and with differences in orientation preference whereas the negative ‘inhibitory’ connections were less selective. Because of its superior performance, this estimator likely provides a more physiologically relevant representation of the functional connectivity in dense recordings than the sample correlation matrix.

## Author Summary

It is now possible to record the spiking activity of hundreds of neurons at the same time. A meaningful statistical description of the collective activity of these neural populations – their ‘functional connectivity’ – is a forefront challenge in neuroscience. We addressed this problem by identifying statistically efficient estimators of correlation matrices of the spiking activity of neural populations. Various underlying processes may reflect differently on the structure of the correlation matrix: Correlations due to common network fluctuations or external inputs are well estimated by low-rank representations, whereas correlations due to linear interactions between specific pairs of neurons are well approximated by their pairwise *partial* correlations. In our data obtained from fast 3D two-photon imaging of calcium signals of large

and dense groups of neurons in mouse visual cortex, the best estimation performance was attained by decomposing the correlation matrix into a sparse network of partial correlations ('interactions') combined with a low-rank component. The inferred interactions were both positive ('excitatory') and negative ('inhibitory') and reflected the spatial organization and orientation preferences of the interacting cells. We propose that the composition of the most efficient among many estimators provides a more informative picture of the functional connectivity than previous analyses of neural correlations.

## Introduction

*Functional connectivity* is a statistical description of *multineuronal* activity patterns not reducible to the response properties of the individual cells. Functional connectivity is commonly expressed as the degree of correlation, synchronization, or other forms of statistical association between neurons or in the form of joint coactivation patterns. Functional connectivity is thought to reflect local synaptic connections, shared inputs from other regions, and endogenous network activity. To remove the confounding effects of the stimulus, the average stimulus responses are subtracted or otherwise accounted for. Although functional connectivity is only a phenomenological description, it can serve as a tool to generate hypotheses about how population activity relates to the anatomical architecture of neural circuits and informs about how information is encoded and processed at the population level.

Pearson correlations between the spiking activity of pairs of neurons, are among the most familiar measures of functional connectivity [1–5]. 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 [1, 2, 6–11]. In addition, noise correlations and correlations in spontaneous activity have been hypothesized to reflect key features of the anatomical connectivity [12]. This interpretation is supported by a series of discoveries of nontrivial relationships between neural correlations and other aspects of circuit organization such as the physical distances between neurons [13, 14], their synaptic connectivity [15], stimulus response similarity [3–5, 15–22], cell types [23], cortical layer specificity [24, 25], progressive changes in development and in learning [26, 27], changes due to sensory stimulation and global brain states [5, 21, 28–31] among others.

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 [32–37]. But multineuronal recordings do provide more information than available from an equivalent number of pairs of cells considered separately. 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 just a few principal components but will affect all correlation coefficients. In contrast, the off-diagonal elements of the precision matrix (inverse of the covariance matrix) constitute scaled partial correlations between neuron pairs. The partial correlation coefficient between two neurons reflects their linear association conditioned on the activity of all the other recorded cells. For this reason, partial correlations have been proposed to more closely reflect the direct effects between components of complex systems and have been used to describe gene interactions [38, 39] and the interactions between brain regions [40, 41].

However, estimation of covariance matrices from large populations presents a number 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 large eigenvalues) [42], and poorly conditioned partial correlations [38]. The *sample correlation matrix*, which is an unbiased estimate, is often used to estimate the true correlation matrix. It has many free parameters making it highly sensitive to sampling noise. As a result, on average, the sample correlation matrix is farther from the true correlation matrix than some structured estimates.

In statistics, the technique of deliberately imposing a structure on an estimate in order to improve

its performance is known as *regularization* [38, 43]. To ‘impose structure’ on an estimate means to bias (‘shrink’) it toward a reduced representation known as the shrinkage target. The selection of the optimal target estimate and the optimal amount of shrinkage can be deduced from the training data either analytically or by cross-validation. Some regularization schemes only perform target selection (*dimensionality reduction, variable selection, or feature selection*) and replace the estimate completely. Others only perform shrinkage toward a single target estimate (*shrinkage estimators*). Yet others do both target selection and shrinkage.

Although regularized covariance matrix estimation is commonplace in other fields such as finance [44], functional genomics [38], and brain imaging [41], surprisingly little work has been done to identity optimal regularization of neural correlation matrices. All prior empirical findings from neural correlations have relied on sample correlations. Improved estimation is not the only or even the main benefit of regularization. When we select the best performing estimator among many regularized estimators, we learn something about the system itself: The structure imposed by the best estimator represents the underlying dependencies in the system more closely than the structures imposed by the other estimators. We can then examine this structure as a more effective description of the functional connectivity in the circuit than the sample correlation matrix.

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, in addition to the sample covariance matrix, we evaluated four regularized covariance matrix estimators that imposed different kinds of structure on the estimate. The estimators are designated as follows:

$C_0$  – sample covariance matrix, the unbiased, unregularized estimator.

$C_{\text{diag}}$  – linear shrinkage toward a diagonal covariance matrix.

$C_{\text{factor}}$  – linear shrinkage toward a low-rank matrix produced by factor analysis.

$C_{\text{sparse}}$  – sparse partial correlations, *i.e.* covariance selection.

$C_{\text{sparse+latent}}$  – sparse partial correlations between the recorded neurons combined with partial correlations with several inferred latent units.

First, we demonstrated that, in simulations with known correlation structures, estimators that imposed a matching structure were generally more efficient than other estimators. We therefore reasoned as follows: The most efficient covariance matrix estimator is likely to provide a better picture of the correlation structure than less efficient estimators (See Discussion).

We therefore performed a cross-validated evaluation to establish which of the four regularized estimators was most efficient for 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. We found that  $C_{\text{sparse+latent}}$  consistently outperformed the other estimators. This estimator revealed a sparse network of partial correlations between the observed neurons and inferred several latent units with partial correlations with the observed cells. We analyzed these networks of partial correlations and found the following: Whereas noise correlations were predominantly positive the inferred partial correlations had a large fraction of negative values possibly reflecting inhibitory circuitry. Moreover, we found that the inferred positive partial correlations exhibited stronger dependencies with respect to physical distance and differences in preferred orientations than noise correlations. In contrast, the inferred negative partial correlations were less selective.

## Results

**Covariance estimation** The covariance matrix is defined as

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

where  $\mathbb{E} [\cdot]$  denotes expectation; the  $p \times 1$  vector  $x$  is a single observation the firing rates of  $p$  neurons over time  $\Delta t$ ; and  $\mu$  is the vector of expected firing rates. The usual way to estimate the covariance matrix is to calculate the *sample covariance matrix*  $C_0$  from the empirical sample of observations  $x(t)$ ,  $t = n(1, \dots, n)$  as

$$C_0 = \frac{1}{\nu} \sum_{t=1}^n (x(t) - \bar{x})(x(t) - \bar{x})^T, \quad \bar{x} = \frac{1}{n} \sum_{t=1}^n x(t) \quad (2)$$

where  $\nu$  is the number of degrees of freedom in one neurons activity within the sample ( $\nu = n - 1$  if observations are independent). In this study, we estimate the true mean,  $\mu$ , using the sample mean  $\bar{x}$ , but seek a better estimate of the covariance matrix than  $C_0$ . In the case of noise correlations, the true mean  $\mu$  and the sample mean  $\bar{x}$  are conditioned on the stimulus.

Given an arbitrary covariance matrix estimate  $C$ , the corresponding correlation matrix  $R$  is calculated by normalizing  $C$  by its diagonal (variance estimates):

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

where  $\circ$  denotes entrywise matrix product (Hadamard product) and  $I$  is the  $p \times p$  identity matrix. Similarly, the matrix of partial correlations  $P$  is computed by normalizing the *precision matrix*  $C^{-1}$ :

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

We considered four regularized estimators based on distinct families of reduced target estimates:  $C_{\text{diag}}$ ,  $C_{\text{factor}}$ ,  $C_{\text{sparse}}$ , and  $C_{\text{sparse+latent}}$ . In probabilistic models with exclusively linear dependencies, the targets estimates of these for estimators can correspond to distinct families of graphical models (Fig. 1 Row 1).

The target estimate of estimator  $C_{\text{diag}}$  is the diagonal matrix  $D$  containing estimates of neurons variances. Regularization is achieved by linear *shrinkage* of the unbiased estimate  $C_0$  toward  $D$  controlled by the scalar *shrinkage intensity* parameter  $\lambda \in [0, 1]$ :

$$C_{\text{diag}} = (1 - \lambda)C_0 + \lambda D \quad (5)$$

The structure of imposed by  $C_{\text{diag}}$  favors populations with no linear associations between the neurons (Fig. 1 Row 1, A). If sample correlations arise from spurious or contingent effects,  $C_{\text{diag}}$  is expected to get closer to truth than other estimators.

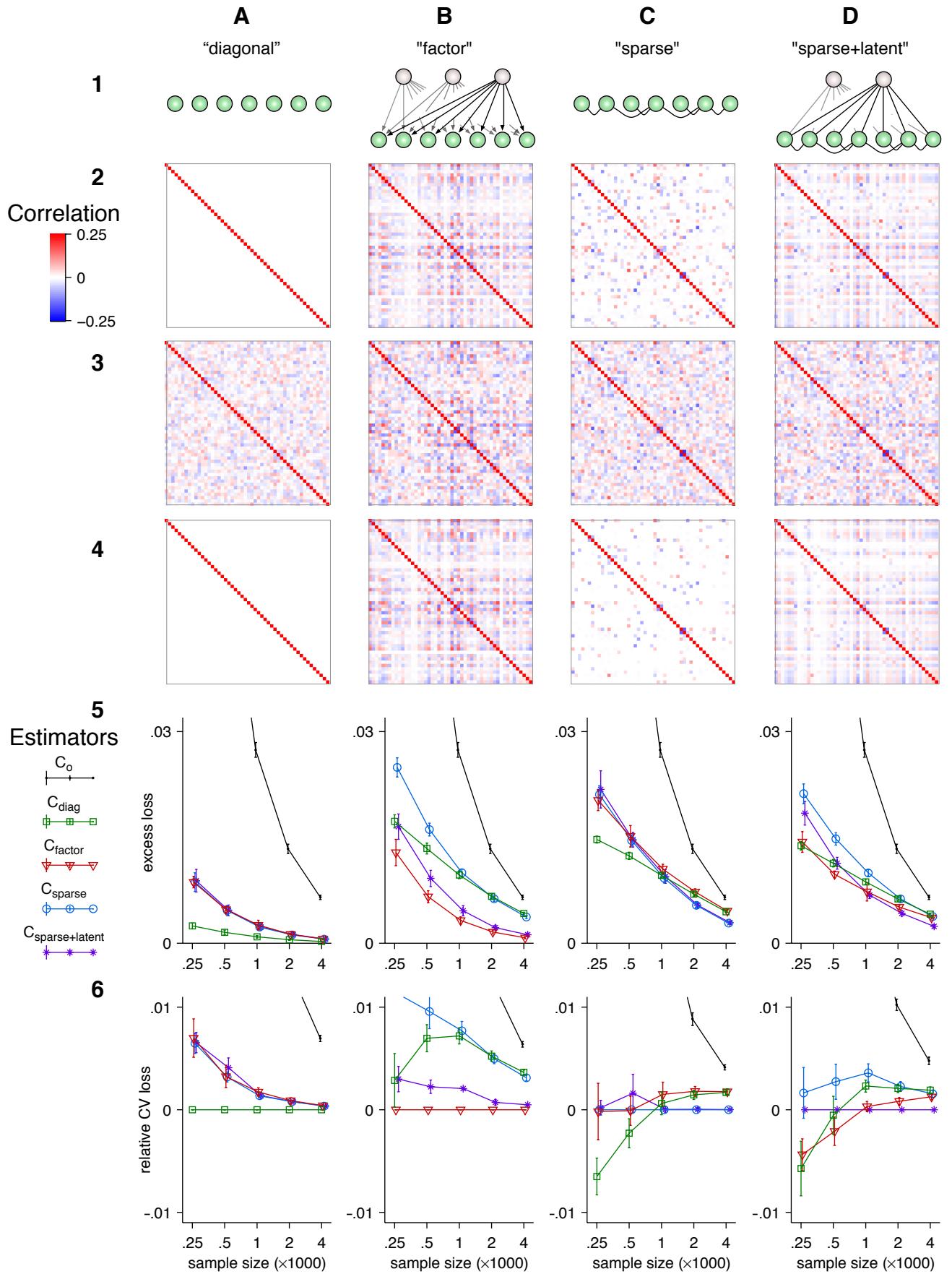
The target of estimator  $C_{\text{factor}}$  is the factor model  $L + \Psi$ , where  $L$  is the  $p \times p$  positive definite matrix with low rank and  $\Psi$  is the diagonal matrix of independent variances. The estimator,

$$C_{\text{factor}} = (1 - \lambda)C_0 + \lambda(L + \Psi), \quad (6)$$

has two hyperparameters: the number of factors in  $L$  and the shrinkage intensity  $\lambda$ . The structure imposed by  $C_{\text{factor}}$  favors conditions in which the population activity is linearly driven by a number of latent factors that affect many cells while direct interactions between cells are insignificant (Fig. 1 Row 1, B).

Estimator  $C_{\text{sparse}}$  is produced by approximating the sample covariance matrix by a matrix whose inverse matrix is sparse:

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



where  $S$  is a sparse matrix with a large fraction of zeros in its off-diagonal. The estimator has one hyperparameter that determines the sparsity (fraction of off-diagonal zeros) in  $S$ . The structure imposed by  $C_{\text{sparse}}$  favors conditions in which neural correlations arise from direct linear effects ('interactions') between some pairs of neurons (Fig. 1 Row 1, C). We will sometimes refer to sparse partial interactions between neurons as simply 'interactions'.

Estimator  $C_{\text{sparse+latent}}$  is produced by approximating the sample covariance matrix by a matrix whose inverse is the sum of a sparse component and a low-rank component:

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

where, as above,  $S$  is a sparse matrix and  $L$  is a low rank matrix. The estimator has two hyperparameters: one to regulate the sparsity of  $S$  and the other to regulate the rank of  $L$ . 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 linear effects from several latent units (Fig. 1 Row 1, D) [45, 46].

**Figure 1 (preceding page).** Estimators whose regularization targets represent the structure of the true covariance matrix outperform other estimators. **Row 1.** Graphical representations of the target estimates used by the four respective regularized covariance matrix estimators. Recorded neurons are denoted as green spheres and latent units as light-shaded spheres. The edges between them denote non-zero partial correlations. In estimator  $C_{\text{diag}}$ , the regularization target is a diagonal matrix. This target corresponds to a model wherein observed neurons exert no linear effects on one another (**Row 1, A**). In estimator  $C_{\text{factor}}$ , the regularization target is a factor model (low-rank matrix plus a diagonal matrix). This target describes a model wherein the activity of observed neurons is linearly driven by several latent units (**Row 1, B**). In estimator  $C_{\text{sparse}}$ , the covariance matrix is approximated as the inverse of a sparse matrix. This approximation describes a model wherein correlations arise due to linear effects exerted by the recorded cells onto one another other (**Row 1, C**). In estimator  $C_{\text{sparse+latent}}$ , the covariance matrix is approximated as the inverse of a matrix constructed as the sum of a sparse and a low-rank component. This approximation describes a model wherein correlations arise due to linear affects exerted by the recorded cells onto one another and due to linear effects from several latent units (**Row 1, D**).

**Row 2:** Examples of  $50 \times 50$  correlation matrices corresponding to each structure: the diagonal correlation matrix (**A**), a factor model with three latent units (**B**), a correlation matrix with 73% off-diagonal zeros (**C**), and a correlation matrix whose inverse is the sum of a rank-1 (*i.e.* one latent unit) component and a sparse component with 78%[1] off-diagonal zeros.

**Row 3:** Sample correlation matrices calculated from samples of 1000 observations drawn from simulated random processes with respective correlation matrices from row 2. The structure of the correlation matrix is much more difficult to discern by eye.

**Row 4:** Regularized correlation matrix estimates computed from the same samples as in Row 3. For each kind of ground truth, the estimator with the matching structure was used and the regularization parameters optimized by cross-validation using the data in the same sample. The regularized estimates are closer to truth than sample correlation matrices.

**Row 5:** Excess losses (Eq. 11) attained by each of the five estimators for various sample sizes. The error bars indicate the standard error of the mean based on 30 samples. Estimators with the matching structure converged to zero faster than the other estimators when the sample size was large enough to allow the more complex true structures outperform simpler structures.

**Row 6:** Relative cross-validation losses (Eq. 13) attained by each of the five estimators with respect to the matching estimator. Error bars indicate the standard error of the mean based on 30 samples.

**Simulation** To demonstrate how the selection between regularized estimators allows discerning the structure of correlation, we constructed four  $50 \times 50$  covariance matrices with the structures imposed by the four respective regularized estimators (Fig. 1 Row 2, A–D). See the Methods section for details of how these matrices were generated. We used these covariance matrices as the ground truth in multivariate normal distributions with zero means and drew samples of various sizes. Even with relatively large sample sizes (*e.g.*  $n = 1000$ ), the sample correlation matrices were contaminated by sampling noise and their underlying structure was difficult to discern (Fig. 1 Row 3).

An evaluation of a covariance matrix estimators requires a *loss function*  $\mathcal{L}(C, \Sigma)$  to quantify the discrepancy between the estimate,  $C$ , and the truth,  $\Sigma$ . The loss function must attain its minimum when  $C = \Sigma$ .

In this study, we adopted the *normal loss* function defined as

$$\mathcal{L}(C, \Sigma) = \frac{1}{p} [\ln \det C + \text{tr}(C^{-1}\Sigma)] \quad (9)$$

This loss function derives from the multivariate normal log-likelihood function  $L(\Sigma|C_0)$  through the following expression:

$$\mathcal{L}(\Sigma, C_0) = -\ln(2\pi) - \frac{2}{p} L(\Sigma|C_0) \quad (10)$$

The normalization by  $p$  makes the values of the loss function comparable across different population sizes.

Although the two functions appear similar in form, they differ conceptually: The log-likelihood  $L(\Sigma|C_0)$  is a function of the parameter  $\Sigma$  given the sample covariance matrix  $C_0$  whereas the loss function expresses the deviation of an arbitrary estimate  $C$  from the truth  $\Sigma$ .

The choice of normal loss function is motivated by mathematical convenience. We expect that the main conclusions of our study will not change qualitatively with other well behaved loss functions, such as Stein's entropy loss or quadratic loss [38, 42, 47, 48].

The *excess loss* function is produced by shifting the loss function to zero at its minimum:

$$\mathcal{L}_0(C, \Sigma) = \mathcal{L}(C, \Sigma) - \mathcal{L}(\Sigma, \Sigma), \quad (11)$$

We drew 30 independent samples from each model of size  $n = 250, 500, 1000, 2000$ , and 4000 and measured the excess loss produced by 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 excess loss) than the sample covariance matrix. However, estimators whose structure matched the true model typically outperformed the other estimators (Fig. 1 row 5). There were two exceptions: First, for some smaller samples, the data were insufficient to reveal the true correlation structure and the estimators with simpler structures,  $C_{\text{diag}}$  and  $C_{\text{factor}}$  often outperformed the estimators with matching structure (Fig. 1 Row 5, C and D). Second,  $C_{\text{sparse+latent}}$  often performed equally well to  $C_{\text{sparse}}$  on the model without latent units. This can be explained by the fact that  $C_{\text{sparse+latent}}$  correctly inferred the right number of latent units, zero.

Even though, with increasing sample sizes, all estimators converged to the truth (zero excess loss), the distinction between the estimators did not decrease.

In empirical data the ground truth,  $\Sigma$ , is not accessible and excess loss cannot be computed directly. Instead, the loss function can be estimated solely from the data through *validation*. In validation, an independent *testing sample* is set aside to compute an additional sample covariance estimate  $C'_0$  to validate the estimate  $C$  computed from the main, *training sample*. Then *validation loss*  $\mathcal{L}(C, C'_0)$  can substitute the loss function  $\mathcal{L}(C, \Sigma)$ .

The normal loss (Eq. 9) is particularly convenient because it is additive in its second argument in the sense that

$$\mathcal{L}(C, z_1) + \mathcal{L}(C, z_2) \equiv \mathcal{L}(C, z_1 + z_2)$$

In this case, the cross-validation loss is an unbiased estimate of the loss function:

$$\mathbb{E}_{C'_0} [\mathcal{L}(C, C'_0)] = \mathcal{L}(C, \mathbb{E}_{C'_0}[C'_0]) = \mathcal{L}(C, \Sigma)$$

The property of additivity does not hold for other popular loss functions such as Stein's entropy loss or various quadratic losses; their validation losses are biased estimates of the loss function.

Acquiring a separate testing sample is not practical. Instead,  $K$ -fold cross-validation can be used: The sample is divided into  $K$  subsets of approximately equal size ( $K = 10$  in this study). Then each subset is used as the validation sample with the other  $K - 1$  serving as the training sample. The validation losses from each of such 'folds' are averaged to produce the *cross-validation loss* or *CV-loss* for short. Let  $C_0^{\{k\}}$  denote the sample covariance matrix computed from the  $k^{th}$  subset and  $C^{\{\setminus k\}}$  denote the results of estimator  $C$  trained on the remaining  $K - 1$  subsets. Then cross-validation loss for estimator  $C$  is

$$\ell_C = \frac{1}{K} \sum_{k=1}^K \mathcal{L}\left(C^{\{\setminus k\}}, C_0^{\{k\}}\right) \quad (12)$$

In this formulation,  $\ell_C$  can be recognized as the cross-validated Gaussian log-likelihood (up to a constant offset and multiplier). However, the present framework is more general, allowing for other loss functions to be used.

Since we only need to compare estimators to each other, we are only interested in the *relative* CV loss of estimator  $C$  with respect to reference estimator  $C_{ref}$ :

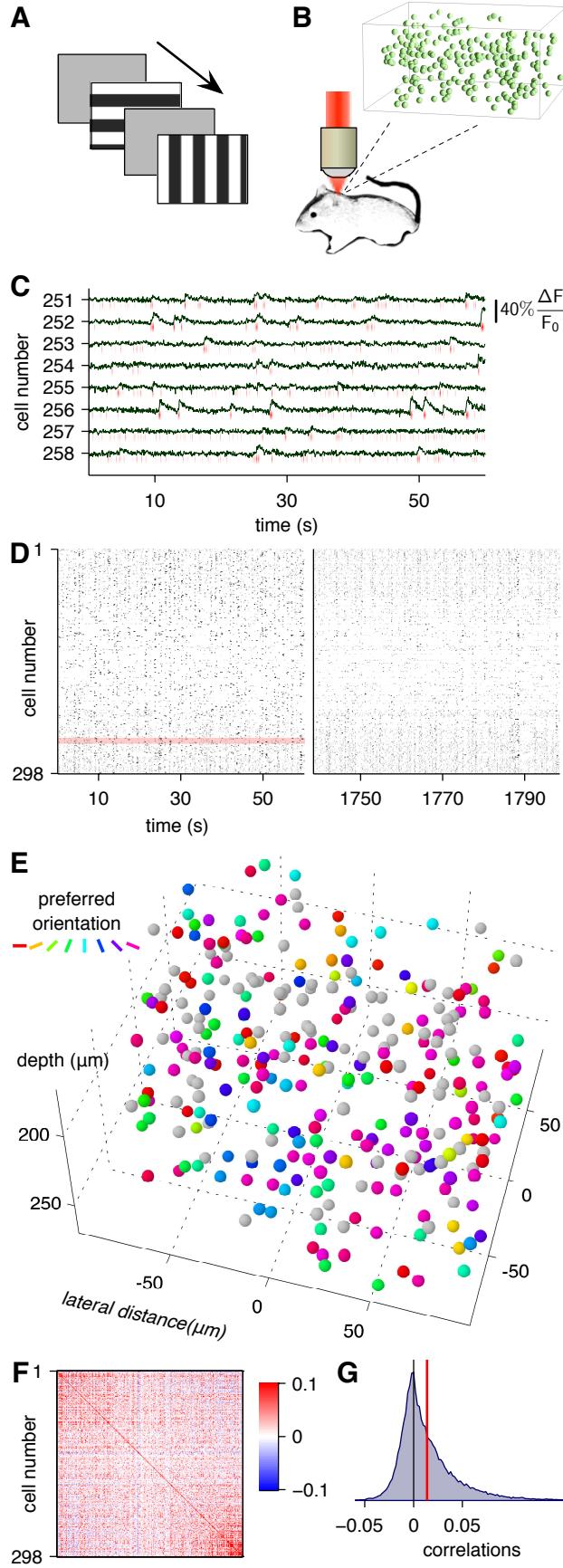
$$\Delta\ell_{C,C_{ref}} = \ell_C - \ell_{C_{ref}} \quad (13)$$

To examine how well the relative CV-losses reproduced the relationships between the excess losses, we evaluated the relative CV-loss with respect to the estimator that matched the structure of the data-generating model. Indeed, CV-losses accurately reproduced the relationships between the excess losses of the estimators, with somewhat greater variability (Fig. 1 Row 6).

This simulations study demonstrated that, with sufficiently large sample sizes, imposing the correct type of structure on the estimate, led to greater estimation improvement than imposing incorrect structures

**Covariance estimation in neural data** We recorded the calcium activity of dense populations of neurons in the supragranular layers in primary visual cortex of anesthetized mice using fast random-access 3D scanning two-photon microscopy during visual stimulation (Fig. 2 A–B) [49–51]. 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. 2 C). The firing rates of neurons were inferred using sparse nonnegative deconvolution [52] to produce (Fig. 2 C and D). Only cells that produced detectable calcium activity were included in the analysis (For details, see Methods). First, full-field gratings moving in 16 directions were presented to determine the orientation tuning of the recorded cells (Fig. 2 E). To estimate the noise correlation matrix, we presented 300[2] repetitions of full-field drifting gratings with two directions of motion in some experiments or 150[3] repetitions with five directions in other experiments. Each grating lasted 1 second and was followed by a 1-second blank. The average stimulus response was subtracted from each trial; the traces were then binned into 150 ms intervals for the estimation of the covariance matrix. The sample correlation coefficients were largely positive and low (0.01–0.03 in most images sites) (Fig. 2 F and G).

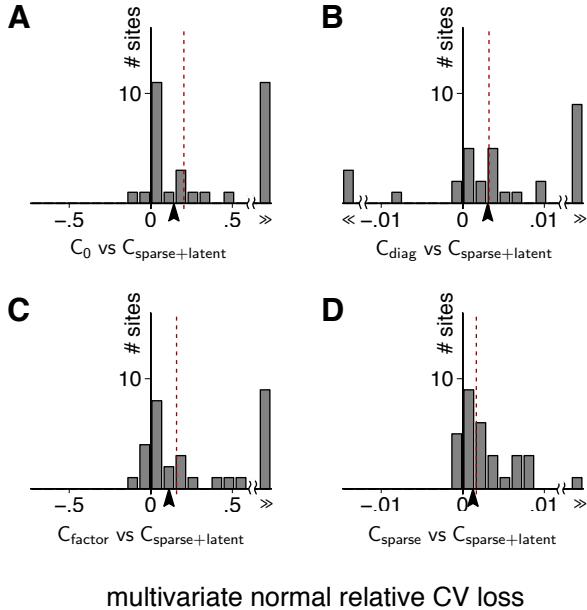
In these dense local populations direct interactions between cells are likely to contribute substantially to the patterns of population. We therefore hypothesized that covariance matrix estimates that explicitly model the partial correlation between pairs of neurons ( $C_{\text{sparse}}$  and  $C_{\text{sparse+latent}}$ ) would have a performance advantage over those that do not. However, the observed neurons must also be strongly influenced by



**Figure 2. Acquisition of neural signals for the estimation of noise correlations.** 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–G. Calcium activity data from an example site. C. Representative calcium signals from eight cells out of 298 cells downsampled to 20 Hz. The inferred firing rate binned in 150 ms intervals are indicated by red ticks below each trace. D. The raster plot of the inferred firing rates, binned in 150 ms intervals, from the entire population from the first (left) and last (right) minute of the entire recording. The traces from C are highlighted in red. E. The spatial arrangement and orientation tuning of the 298 cells from the imaged site. F. The noise correlation matrix of the activity of the neural population. G. Histogram of noise correlation coefficients. The red line indicates the mean.

global activity fluctuations and by unobserved common inputs, giving an advantage to estimators that explicitly model common fluctuations of the entire population  $C_{\text{factor}}$  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 estimated the relative CV loss (Eq. 13) of estimators  $C_0$ ,  $C_{\text{diag}}$ ,  $C_{\text{factor}}$ , and  $C_{\text{sparse}}$  with respect to  $C_{\text{sparse+latent}}$ . The hyperparameters of each estimator were optimized by nested cross-validation (See Methods for details). Indeed, the median relative CV loss was significantly above zero (Fig. 3), indicating that  $C_{\text{sparse+latent}}$  produced estimates that were consistently closest to the true covariance matrix.



**Figure 3. The sparse+latent estimator  $C_{\text{sparse+latent}}$  outperforms the other estimators on neural data.** **A–D.** Histograms of average cross-validation loss differences of the respective estimators  $C_0$ ,  $C_{\text{diag}}$ ,  $C_{\text{factor}}$ , and  $C_{\text{sparse}}$  from  $C_{\text{sparse+latent}}$ . The histograms are based on 31 imaged sites in 24 mice. All medians (red dashed lines) were significantly greater than zero, indicating the dominance of  $C_{\text{sparse+latent}}$  over the other estimators. The arrow heads indicate the results for the site shown in Fig. 2 and Fig. 4.

- A.**  $C_{\text{sparse+latent}}$  outperforms  $C_0$ : median improvement 0.16 nats/neuron,  $p = 3.9 \times 10^{-5}$ .
- B.**  $C_{\text{sparse+latent}}$  outperforms  $C_{\text{diag}}$ : median improvement 0.0032 nats/neuron,  $p = 2.2 \times 10^{-3}$ .
- C.**  $C_{\text{sparse+latent}}$  outperforms  $C_{\text{factor}}$ : median improvement 0.16 nats/neuron,  $p = 6.9 \times 10^{-5}$ .
- D.**  $C_{\text{sparse+latent}}$  outperforms  $C_{\text{sparse}}$ : median improvement 0.0016 nats/neuron,  $p = 5.5 \times 10^{-6}$ .

Furthermore,  $C_{\text{sparse}}$  outperformed all other estimators except  $C_{\text{sparse+latent}}$  (Fig. S1) suggesting the greater importance of modeling interactions between pairs of neurons than the common fluctuations of activity across the entire population.

Evaluation with respect to the quadratic loss function,

$$\mathcal{L}(C, C'_0) = \frac{1}{p^2} \text{tr}(C^{-1}C'_0 - I)^2, \quad (14)$$

produced the same relationship between the estimators as the normal loss (Fig. S2) suggesting that the results of this study do not depend on the assumption of gaussianity and reflect real regularities in the data.

**Correlation structure and circuit architecture** Because of its superior estimation performance,  $C_{\text{sparse+latent}}$  likely provides a more physiologically relevant representation of the functional connectivity than the other estimators. The estimator segregates two sources of correlations: linear dependencies between pairs of neurons (sparse component) suggestive of direct interactions and common fluctuations across the entire population (low-rank) component suggestive of common inputs, cell assembly activation, network fluctuations or other forms of synchronized activations.

We examined the structure of the  $C_{\text{sparse+latent}}$  estimate in each imaged site. The regularized estimate of the correlation matrix appeared very similar to the sample correlation matrix (Fig. 4 A and D) but the corresponding partial correlation matrices differed substantially. The partial correlation matrix (Fig. 4 B

and E) and a low-rank component (Fig. 4 C). While sample correlations were mostly positive, the sparse partial correlations had a high fraction of negative interactions. In the example site (Fig. 2 and Fig. 4)), the sparse component had 82.1% [4]sparsity (or, conversely, 17.9% [5]connectivity) with average node degree of 52.5[6] (Fig. 4 G). The low-rank component had rank 17.

We compared the graphs of partial correlations revealed by  $C_{\text{sparse+latent}}$  to the graph obtained by thresholding the sample correlations. Thresholded correlations have been used in several studies as to isolate the likely pairwise interactions with subsequent graph theoretical analysis [53]. We set the threshold to match the connectivity of the  $C_{\text{sparse+latent}}$  graph. The networks revealed by the two methods differed substantially. In the example site with 17.9%[11] connectivity, only 42%[12] of cell pairs connected in the  $C_{\text{sparse+latent}}$  graph were also connected in the thresholded correlation graph (Fig. 4 F, H, and I). In particular, the thresholded sample correlations were predominantly positive whereas 38%[13] of the interactions in  $C_{\text{sparse+latent}}$  were negative. Most of the negative interactions corresponded to low sample correlations (Fig. 4 F). Thus the  $C_{\text{sparse+latent}}$  estimates reveal negative interactions that account for low correlations where high correlations are predicted from all the other interactions.

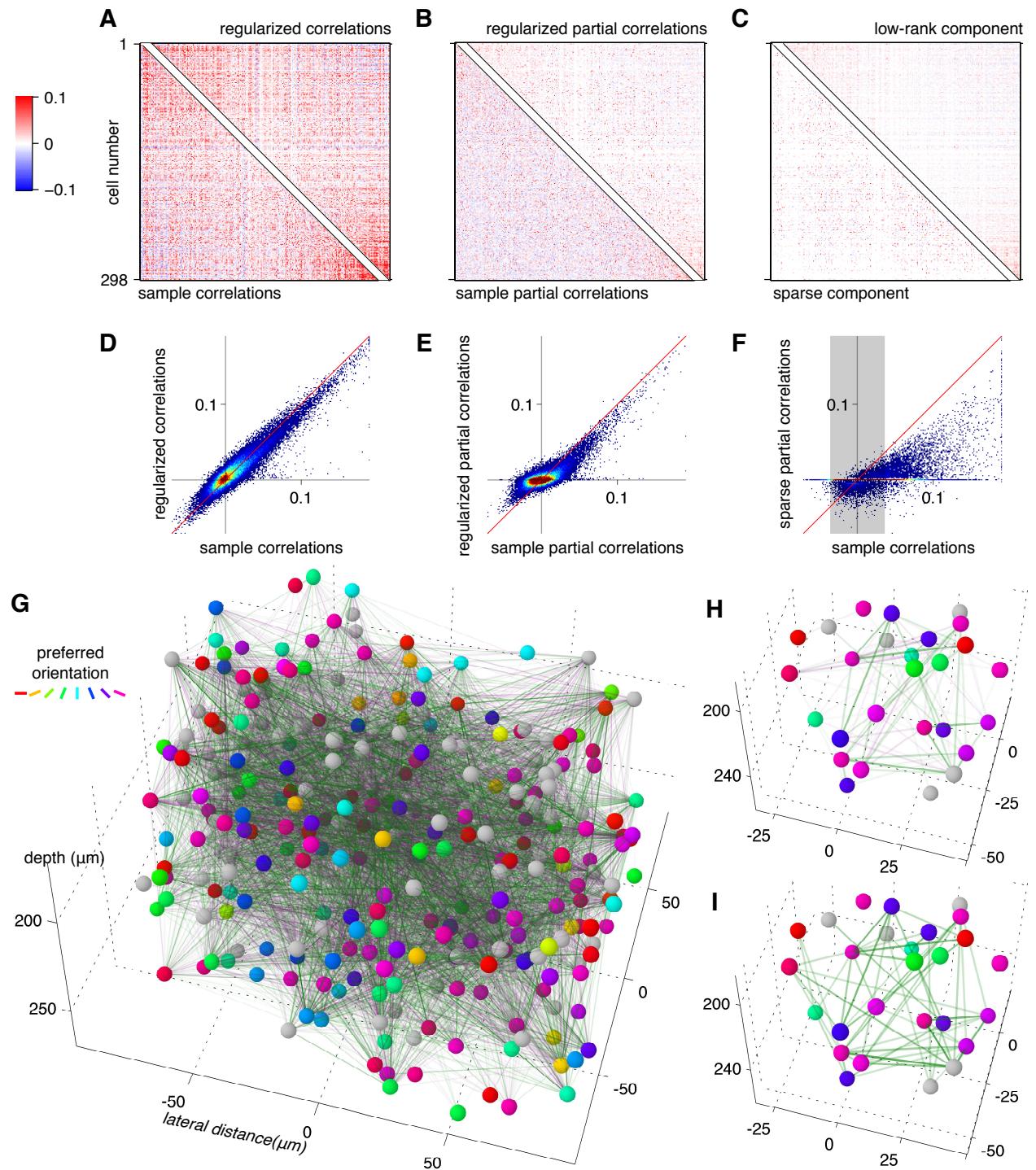
The average partial correlations revealed by  $C_{\text{sparse+latent}}$  estimates in all 31 sites were about 5 times lower than the sample correlations and less variable across sites (Fig. 5 A): the coefficient of variation of the average sample correlations across all sites was 0.41 whereas that of the average partial correlations was only 0.35. The average node degree of the sparse component of the partial correlations and the number of inferred latent units varied widely between sites but generally increased with recorded population size (Fig. 5 B and C). However, there was an inverse relationship between the number of latent units and the average node degree (Fig. 5 D). Several sites, even with relatively large population sizes, had fairly few pairwise interactions and were dominated by latent units. These differences have multiple explanations, including differences in brain states and recording quality (neuropil contamination, motion).

We also examined the relationship of the correlation structure to the differences in orientation preference and to the geometric distances between the cell pairs (Fig. 6). The partial correlations of the  $C_{\text{sparse+latent}}$  estimates fell more rapidly with the difference in preferred orientation (Fig. 6 A). Similarly, the regularized partial correlations decreased more rapidly than sample correlations with geometric distance laterally (Fig. 6 B) and in depth (Fig. 6 B).

Furthermore, the positive and negative interactions in the  $C_{\text{sparse+latent}}$  estimator were organized differently: The positive interactions fell rapidly with differences in the preferred orientation (Fig. 6 D), lateral displacements (Fig. 6 E), and displacement in depth (Fig. 6 F) while the negative interactions were much less selective (Fig. 6 D–F).

## Discussion

**Functional connectivity as a network of pairwise interactions** Functional connectivity is often expressed as a graph of pairwise interactions. The intended purpose of many studies of functional connectivity is to estimate the circuits anatomical connectivity solely from the 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 or shared synaptic inputs [12, 14, 32, 33, 54]. Such signatures are still ambiguous as they can arise from network effects other than direct synaptic connections [55]. With simultaneous recordings from a larger subset of neurons in the circuit, 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 rest of the network. Therefore, conditionally independent neurons can be inferred to lack a direct causal influence. Such conditional dependencies can be inferred by fitting a probabilistic model of the joint population activity. For example, generalized linear models (GLMs) can include biophysically plausible synaptic integration and membrane kinetics, individual neurons stimulus drive, and a network of synaptic couplings [56]. Maximum entropy models are similarly constrained by



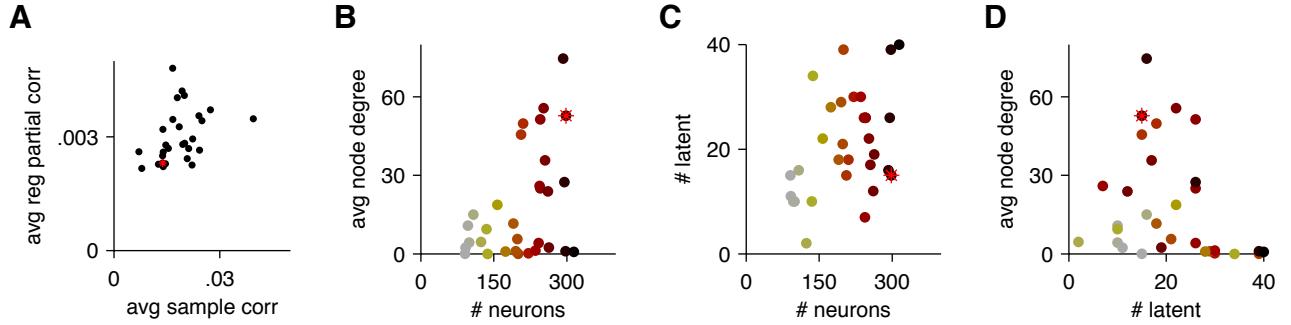
observed pairwise correlations [57–61]. Although not informed by physiology, maximum entropy models are attractive because they account for the observed statistics with minimum other assumptions about the distribution: the investigator need only choose which statistics to match. Under the multivariate 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 different network of ‘interactions’, *i.e.* conditional dependencies between pairs of neurons.

Inference of the conditional dependencies also depends on the completeness of the recorded population: To properly condition the interaction between a pair of neurons on the activity of the other neurons under a given probabilistic model, the recording must include all the other neurons with which the pair can potentially interact. Otherwise, interactions with unobserved portions of the circuit can introduce false conditional dependencies between observed neurons. For this reason, the most successful applications of statistical models of population activity have been in *in vitro* preparations of the retina or cell cultures where high-quality recordings from the complete populations have been possible [56]. In cortical tissue, electrode arrays record from a small fraction of cells in a given volume, limiting the validity of the inference of the conditional dependencies between pairs of cells. Perhaps for this reason, partial correlations have not been previously used to describe the functional connectivity between cortical neurons. Despite this limitation, many studies have fitted probabilistic models to multielectrode recordings and analyzed the structure of the resulting conditional dependencies [59, 60].

Correct inference of causal interactions between neurons depends on correct mathematical representations of the interactions in the statistical model and the completeness of the recorded population. It is not yet clear which approach provides best correspondence to the anatomical connectivity in the circuit, with little experimental evidence yet available to address this question. The connectivity graphs inferred by various statistical methods are commonly reported as abstract representations of the functional connectivity without a statement about its anatomical implications. Topological properties of such graphs have been interpreted to indicate principles of circuit organization such as small-world organization [53, 59, 62]. However, the topological properties of graphs depend critically on the method of their inference (Zalesky:2012) and, until the physiological meaning of the functional connectivity is confirmed, such analyses may be questionable.

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 advanced spike inference algorithms provide millisecond-scale temporal resolution of single action potentials [63]. However, such high temporal precision comes at

**Figure 4 (preceding page). Example of low-dimensional correlation structure revealed by the sparse+low-rank estimator.** **A.** The regularized estimate of the correlation matrix (top-right) closely approximates the sample correlation matrix (bottom left). This close approximation is also demonstrated by the scatter plot of the correlation coefficients produced by the two estimates (**D**). However, the partial correlation matrices from the two estimates show more pronounced differences (**B** and **E**). **C.** The partial correlation matrix of the regularized estimate is decomposed into a sparse component with 82.1%[7] off-diagonal zeros (bottom-left) and low-rank component of rank 17[8] (top-right). **F.** The sparse component of the regularized partial correlation matrix had little resemblance to the sample correlations. The gray interval indicates the range of correlations containing 82.1%[9] of cells pairs, equal to the fraction of zeros in the sparse partial correlation matrix. This interval contained 58.9%[10] of the partial correlations. **G.** A graphical depiction of the positive (green) and negative (magenta) partial correlations as edges between observed neurons. The line density is proportional to the magnitude of the correlation. **H.** A subset of neurons from the center of the cluster shown in **G** showing the regularized partial correlations. **I.** The same subset with sample correlations thresholded to match the sparsity of the regularized interactions.



**Figure 5. Properties of sparse+low-rank regularized estimates from all imaged sites**

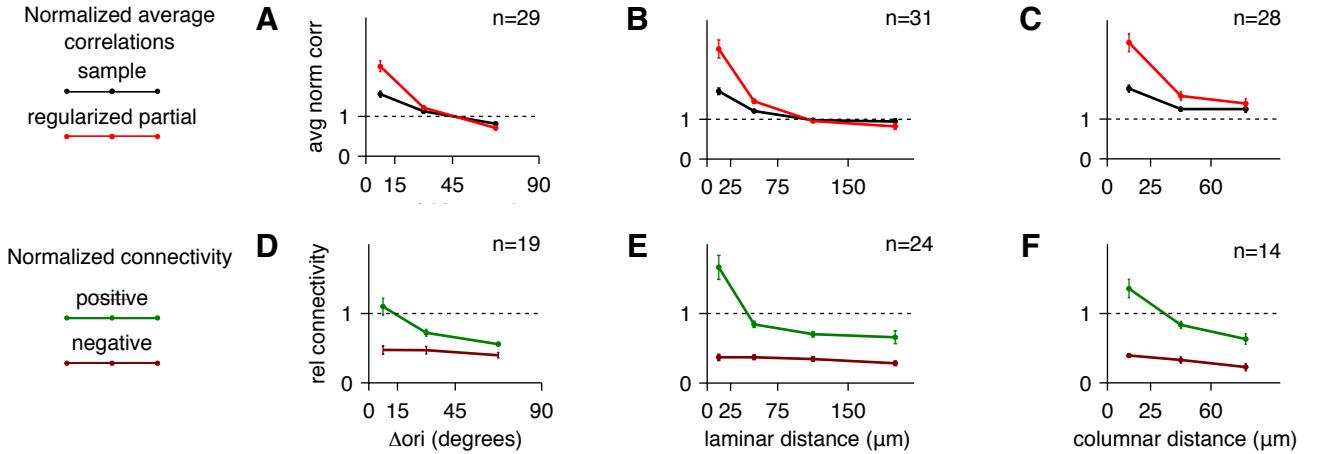
**A.** The average sample correlations vs. average partial correlations for each imaged site. In each plot, the red asterisk indicates the site shown in figures 2 and 4. **B.** The average node degree for sparse partial correlations vs. population size in each imaged site. **C.** The number of inferred latent units vs. population size in each imaged site. **D.** The number latent units vs. average node degree for sparse partial correlations for each site.

the cost of the accuracy of the inferred spike rates: better accuracy can be achieved when calcium signals are analyzed on the temporal scales of tens of milliseconds [51]. 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* [50,51] and even from entire nervous systems [64,65]. These techniques may open new opportunities for more incisive of functional connectivity than with 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 maximum entropy models on the binary domain). Most studies of functional connectivity have relied on instantaneous sample correlations [23, 26, 28, 53]. 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 a statement about underlying mechanisms.

In our study, we sought to infer pairwise functional connectivity networks among neurons in cortical microcircuits. We hypothesized that partial correlations would correspond more closely to underlying mechanisms than sample correlations. Since neurons form synaptic connections mostly locally and sparsely [66], we a priori favored solutions with sparse partial correlations. The importance of partial correlations may be justified by the principle of maximum entropy: The maximum entropy distribution on discrete or continuous multivariate domains constrained by the observed mean firing rates, their variances, and correlations is the multivariate normal distribution. The interactions (conditional dependencies) between pairs of neurons inferred by this distributed are expressed by the precision matrix or, equivalently, by the matrix of the pairwise partial correlations. Therefore, under the aforementioned 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 describe the functional connectivity of a circuit is to isolate patterns of multineuronal coactivations [67–72]. 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 [67,70], *i.e.* groups of tightly interconnected groups



**Figure 6. Dependence of correlations and partial correlations on orientation tuning differences and physical distance between cell pairs.** A–C. Average partial correlations (red) estimated by  $C_{\text{sparse+latent}}$  and average sample correlations (black) averaged across multiple imaged sites. In each site, the correlations were normalized by the respective average correlation shown in Fig. 5 A. The number  $n$  of sites that qualified to be included in the analysis is indicated. Sites were included if they had at least 20 pairs of neurons in each of the intervals. The error bars indicate the standard error of the mean for based on  $n$ . **A.** Average correlations between pairs of neurons tuned to orientation with differences in preferred orientation in the intervals of 0–15°, 15–45° and 45–90°. **B.** Average correlations between pairs of neurons located at the same depth ( $\pm 25\mu\text{m}$ ) separated by lateral distances in the intervals of 0–25  $\mu\text{m}$ , 25–75  $\mu\text{m}$ , 75–150  $\mu\text{m}$ , and 150+  $\mu\text{m}$ . **C.** Average correlations between pairs of neurons displaced laterally by less than 25  $\mu\text{m}$  separated in depth by distances in the intervals of 0–25  $\mu\text{m}$ , 25–60  $\mu\text{m}$ , and 60+  $\mu\text{m}$ . **D–F.** Normalized connectivity of positive (green) and negative (dark red) interactions from the sparse component obtained from  $C_{\text{sparse+latent}}$ . Normalized connectivity was computed as the fraction of pairs connected by interactions of corresponding signs in each interval divided by the fraction of non-zero interactions across the entire site. Site were included in the analysis only if they had at least 20 cell pairs in each interval. The error bars indicate the standard error of the mean based on the number  $n$  of sites included in the analysis.

of cells involved in a common computation. Coactivation patterns could also be explained as shared input from unobserved parts of the circuit. Coactivations could also be explained by global network fluctuations modulating the activity of the local circuit [73].

Coactivation patterns and pairwise connectivity are not mutually exclusive since assemblies arise from patterns of synaptic connectivity, but analysis of coactivations method shifts the focus from detailed interactions to the collective behavior.

In our study, the analysis of functional connectivity through modes of coactivations was represented by the factor analysis estimator  $C_{\text{factor}}$ .

**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 [74], high-order correlations [75], or global network fluctuations [76].

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 [45, 46]. 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 increasingly worse as the size of the recorded population of neurons increases until such models lose their statistical validity [77]. As the recorded neuronal population sizes increased gradually, experimental neuroscientists addressed this problem by extending the recording durations to keep sampling noise in check and verified that existing models are not overfitted and that the stationarity assumption are observed [76]. However, with ambitious projects, such as the BRAIN initiative currently underway [78] aiming to record from significantly larger populations, the problem must be addressed by regularizing the solution. Regularization limits the solution to a smaller subspace to counteract the effect of the sampling noise in the empirical data. However, limiting the solution to an inappropriate subspace does not allow significant improvement.

Several strategies have been developed to limit the model space in order to improve the quality of the estimate. For example, [75] developed a heuristic rule to identify the most significant features that must be fitted by a maximum entropy model for improved performance in retina. Generalized linear models typically employ  $L_1$  penalty terms to constrain the solution space and to effectively reduce the dimensionality of the solution [56].

In our study, regularization was also accomplished by dimensionality reduction (feature selection) schemes to produce sparse, constrained solutions.

Since optimal regularization schemes are specific to systems under investigation, the inference of functional connectivity in big neural data will entail the search for optimal regularizations that may involve combinations of heuristic rules and numerical techniques specifically designed for each type of neural circuit.

**Model selection** In our study, the covariance matrix estimators were evaluated with respect to the cross-validated normal log likelihood. However, this does not limit the applicability of its conclusions to normal distributions. Indeed the major findings in this paper could be reproduced with respect to other loss functions (compare Fig. 3 and Fig. S2). Other probabilistic models, fitted to the same data, would 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, a simple model can produce a better estimate even if it does not correctly represent the real dependencies in the data than a more complex model with correct representation of dependencies. Therefore, showing that a more constrained model has better cross-validated performance than a more complex model does not support the inference that it reveals a better representation of dependencies in the data. This caveat is related to the phenomenon known as *Steins Paradox* [79]: in which the biasing of an estimate toward an arbitrary target is guaranteed to produce *some* improvement over the best unbiased estimate.

**Physiological interpretation and future directions** Although the network of interactions inferred by the optimal estimator,  $C_{\text{sparse+latent}}$ , resembles a network of synaptic interactions, it is unlikely that the anatomical and functional networks are identical. The improved estimation performance gives the model extra credence over other estimates of the functional connectivity. The inferred interaction network was substantially different from that inferred from by considering the network of most significant sample correlations Fig. 4 F, H, and I. The  $C_{\text{sparse+latent}}$  estimator reveals a large number of negative interactions that are not present in the sample correlation matrix (Fig. 4 F). These negative interactions are produced

as explanations for low sample correlation values where high values are expected based on common correlations with other neurons. The sample correlation matrix does not draw attention to such surprisingly low correlations and potentially misses their physiological significance. The resulting network had a larger fraction of negative interactions than in the sample correlation matrix. The negative interactions had a distinct spatial and functional organization from the positive interactions. To determine potential physiological significance of the negative interactions, future experiments with labeled cell types could indicate whether the negative interactions emanate preferentially from inhibitory cells.

The estimator effectively decomposed the correlation structure into two distinct components: pairwise interactions and common coactivation modes denoted as ‘latent units’. Since the distinction led to improved estimation, the two types of components could be further investigated. For example, slow-wave network fluctuations due to anesthesia or slow-wave sleep could be strongly reflected on the common components/latent units but much less on the pairwise interactions. Previous studies found that correlations measured under anesthesia retained little in common to correlations measured in wakefulness and concluded that studies of the anesthetized brain did not inform about its behaviorally relevant functions [28]. When conditioned on the brain state, optimal estimation of covariance matrices in dense cortical networks could reveal that slow-wave activity alters the common components but does not affect the pairwise interactions as strongly. The correlation structure also changes as a function of the stimulus condition [51] on time scales that are much faster than changes in the anatomical connectivity. Descriptions of specific changes in the functional connectivity due to stimulus properties could shed light on how circuits orchestrate their activity to represent diverse stimuli.

We propose that the structure of the most efficient among many estimators provides a better picture of the functional connectivity than that obtained from unbiased sample correlations.

## Methods

**Surgery and two-photon Imaging** 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. The surgical procedures and data acquisition were performed as described in [51]. Briefly, C57BL/6J mice (aged p40–60) were used. Anesthesia was initiated with isoflurane (3%) and the 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 area. 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 the 3D-RAMP two-photon microscope described in [51]. First, A 3D stack was acquired and cells manually segmented. To collect calcium signals, the system repeated hopped between the selected neurons.

**Visual stimulus** Full-field drifting gratings with 90% contrast, luminance of 10 cd/m<sup>2</sup>, spatial frequency of 0.08 cycles/degree, and temporal frequency of 2 cycles/s. Two sets of stimuli were presented for each imaging site: the first to map directional tuning and the second to estimate noise correlations. Directional tuning was mapped using a pseudo-random sequence of drifting gratings at sixteen equally spaced directions of motion changing at 2 Hz for 3 min without blanks. The data for covariance estimation were collected during presentations of full-field drifting gratings with the same parameters as those used in directional tuning except only two directions (in 9 datasets) or five directions (in 22 datasets) were used and the presentations lasted 1 second and, separated by 1-second blanks. Each stimulus condition was presented at least 180 times.

**Data processing** All data were processed in MATLAB using the DataJoint data processing chain toolbox ([datajoint.github.com](https://datajoint.github.com)) first developed in our lab.

The collected fluorescent traces were deconvolved to reconstruct the firing rates for each neuron. First, the first principal component was subtracted from the traces, which reduced common mode noise related to small movement and cardiovascular artifacts. The resulting traces were low-pass filtered below 0.1 Hz and downsampled to 20 Hz. Firing rates were estimated using a fast non-negative deconvolution algorithm [52].

Orientation tuning was computed by fitting the mean firing rates in response to gratings of directions  $\phi$  with two-peaked von Mises tuning functions of the form  $f(\phi) = a + b \exp\left[\frac{1}{w}(\cos(\phi - \theta) - 1)\right] + c \exp\left[\frac{1}{w}(\cos(\phi - \theta + \pi) - 1)\right]$  where  $b \geq c$  are amplitudes of the two respective peaks,  $w$  is the tuning width, and  $\theta$  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  $p$ -value of the fit was computed as the fraction of permuted datasets for which the  $R^2$  value of the tuning function fit exceeded that of the real data. Cells were considered tuned for  $p$ -values not exceeding 0.05.

For covariance estimation, the analysis was limited to the period with 2 or 5 stimulus conditions and lasted between 12 and 20 minutes[14]. Only spikes trains whose variance was greater than 1%[15] of the median across all the traces in the recording were included in the analysis.

**Cross-validation** To compare the performance of the estimator against each other, we used conventional 10-fold cross validation to measure cross-validation loss (Eq. 12). Briefly, each recording was split into 30 blocks of equal duration. The 30 blocks were then grouped randomly into 10 datasets with 3 blocks in each. This procedure ensured sufficient independence between the 10 datasets while still ensuring that each dataset included data from different parts of the recording. Then, each dataset was used as the testing dataset with the rest of the data used for estimating the covariance matrix.

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 optimal values of the hyperparameters. The optimization of the hyperparameters was performed within the inner loop in two phases: random search to find a good starting point and 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.

When cross-validation loss was not required, only the inner loop of cross-validation was used, applied to the entire dataset. This approach was used to compute the covariance matrix estimates and their excess-loss in the simulation study (Fig. 1 rows 3 and 4) and to analyze the partial correlation structure of the sparse+latent estimator (Fig. 4–6).

**Covariance Estimation** Within the inner loop of cross-validation, covariance matrix estimation was performed with fixed hyperparameter values provided by the search algorithm. The computation of each regularized estimator only required the sample covariance matrix  $C_0$  of the training dataset.

Estimator  $C_{\text{diag}}$  (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  $C_0$ ) were shrunk toward (linearly mixed with) their mean value:

$$D = (1 - \alpha)C_0 \circ I + \alpha \frac{1}{p} \text{tr}(C_0)I \quad (15)$$

Then the diagonal matrix  $D$  was used as the target of covariance shrinkage (Eq. 5) to produce the final regularized estimate.

The `corpcor` package in the programming language R also implements this estimator [80], although its analytical optimization of the shrinkage intensities is based on the mean squared error whereas we optimized them with respect to the loss function in Eq. 9.

Estimator  $C_{\text{factor}}$  used two hyperparameters: the number of latent factors  $d$  and the shrinkage intensity  $\lambda \in [0, 1]$ . The  $p \times d$  factor loading matrix  $L$  and individual variances  $\Psi$  were computed by solving the

minimization problem

$$(L, \Psi) = \arg \min_{\tilde{L}, \tilde{\Psi}} \mathcal{L}(\tilde{L} + \tilde{\Psi}, C_0), \quad (16)$$

which we solved by an expectation-maximization (EM) algorithm. Under our chosen loss function (Eq. 9), this is equivalent to maximum likelihood estimation of  $L$  and  $\Psi$  under the multivariate Gaussian distribution. The final regularized estimate is obtained by linear shrinkage toward the factor model (Eq. 6).

Estimator  $C_{\text{sparse}}$  has one hyperparameter  $\lambda$ , which regulates the sparsity of its precision matrix  $S$ . The precision matrix is computed in two steps: First, the zero structure  $Z$  is determined by minimizing the  $L_1$ -penalized loss.

$$Z = \arg \min_{\tilde{Z} \succ 0} \mathcal{L}(\tilde{Z}^{-1}, C_0) + \lambda \|\tilde{Z}\|_1 \quad (17)$$

where  $\tilde{Z} \succ 0$  denotes the constraint that  $\tilde{Z}$  be a positive definite matrix and  $\|\tilde{Z}\|_1$  is the element-wise  $L_1$  norm of the matrix  $\tilde{Z}$ . This problem formulation is known as *graphical lasso* [81, 82]. To solve this minimization problem, we modified the alternative-direction method of multipliers (ADMM) algorithm developed by [46]. Then, after the zero structure was determined, the remaining coefficients were fitted without penalty:

$$S = \arg \min_{\tilde{S} \in Z^\sharp} \mathcal{L}(\tilde{S}, C_0), \quad (18)$$

where  $Z^\sharp$  denotes the set of positive-definite matrices with zeros in all entries where entries of  $Z$  equal zero. This step was also solved by ADMM. Then the final estimate is the inverse of  $S$  (Eq. 7). Unlike  $C_{\text{diag}}$  and  $C_{\text{factor}}$ , this estimator does not include linear shrinkage: the selection of the sparsity level provides sufficient flexibility to fine tune the regularization level.

Estimator  $C_{\text{sparse+latent}}$  has two hyperparameters: the number of latent units  $d$  and the sparsity level  $\lambda$ . It estimates the larger sparse precision matrix  $S^*$  of the joint distribution of the  $p$  observed neurons and  $d$  latent units.

$$S^* = \begin{pmatrix} S & S_{12} \\ S_{12}^\top & S_{22} \end{pmatrix}, \quad (19)$$

where the  $p \times p$  partition  $S$  corresponds to the visible units and expresses their partial correlation structure, and  $S_{12}$  and  $S_{22}$  are of size  $p \times d$  and  $d \times d$ , respectively. Then the covariance matrix of the observed population is

$$C_{\text{sparse+latent}} = (S^*)^{-1} = (S - S_{12}S_{22}^{-1}S_{12}^\top)^{-1} \quad (20)$$

The matrix  $S_{12}S_{22}^{-1}S_{12}^\top$  has rank  $d$ . Rather than searching for the optimal sparse structure of  $S_{12}$  and  $S_{22}$ , an ill-posed problem, we estimated these components together as one positive-definite  $p \times p$  matrix  $L$  of rank  $d$ .

The estimate is found in two steps. First, we use the ADMM algorithm to find the zero structure of  $S$  by minimizing the  $L_1$ -penalized loss [45, 46]:

$$(Z, \cdot) = \arg \min_{\tilde{Z}, \tilde{L}} \mathcal{L}(\tilde{Z} - \tilde{L}, C_0) + \|\tilde{Z}\|_1 \quad (21)$$

Then we find the sparse and low-rank components of the precision matrix by minimizing the loss

$$(S, L) = \arg \min_{\tilde{S} \in Z^\sharp, \tilde{L}} \mathcal{L}(\tilde{S} - \tilde{L}, C_0) \quad (22)$$

The partial correlation matrix computed from the overall estimate, computed from  $C_{\text{sparse+latent}}$  according to Eq. 4, includes the effects of interactions between the visible and latent units. This estimate of the partial correlations was used in analyses where sparsity was not required (Fig. 4B and E and

Fig. 5 A, and Fig. 6 A–C). For analyses that required a sparse network (Fig. 4 F, G, H and Fig. 5 B, D and Fig. 6 D–F) was computed using the extended precision matrix  $S^*$ :

$$P_{\text{sparse}} = (S \circ I)^{-\frac{1}{2}} S (S \circ I)^{-\frac{1}{2}} \quad (23)$$

The MATLAB code for these computations is available at <http://github.com/atlab/cov-est>.

**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 generated by multivariate normal distributions with the respective true covariance matrices and estimated by each of the estimators.

## Acknowledgments

We thank Genevera Allen for a helpful discussion. We thank Eftychios Pnevmatikakis for helpful suggestions and feedback.

## References

1. Averbeck BB, Latham PE, Pouget A (2006) Neural correlations, population coding and computation. *Nat Rev Neurosci* 7: 358–366.
2. Zohary E, Shadlen MN, Newsome WT (1994) Correlated neuronal discharge rate and its implications for psychophysical performance. *Nature* 370: 140–143.
3. Kohn A, Smith MA (2005) Stimulus dependence of neuronal correlation in primary visual cortex of the macaque. *J Neurosci* 25: 3661–73.
4. Bair W, Zohary E, Newsome WT (2001) Correlated firing in macaque visual area mt: time scales and relationship to behavior. *The journal of Neuroscience* 21: 1676–1697.
5. Ecker AS, Berens P, Keliris GA, Bethge M, Logothetis NK, et al. (2010) Decorrelated neuronal firing in cortical microcircuits. *Science* 327: 584–587.
6. Abbott L, Dayan P (1999) The effect of correlated variability on the accuracy of a population code. *Neural computation* 11: 91–101.
7. Sompolinsky H, Yoon H, Kang K, Shamir M (2001) Population coding in neuronal systems with correlated noise. *Physical Review E* 64: 051904.
8. Nirenberg S, Latham PE (2003) Decoding neuronal spike trains: How important are correlations? *Proceedings of the National Academy of Sciences* 100: 7348–7353.
9. Josic K, Shea-Brown E, Doiron B, de la Rocha J (2009) Stimulus-dependent correlations and population codes. *Neural computation* 21: 2774–2804.
10. Berens P, Ecker AS, Gerwinn S, Tolias AS, Bethge M (2011) Reassessing optimal neural population codes with neurometric functions. *Proc Natl Acad Sci U S A* 108: 4423–8.
11. Ecker AS, Berens P, Tolias AS, Bethge M (2011) The effect of noise correlations in populations of diversely tuned neurons. *The Journal of Neuroscience* 31: 14272–14283.

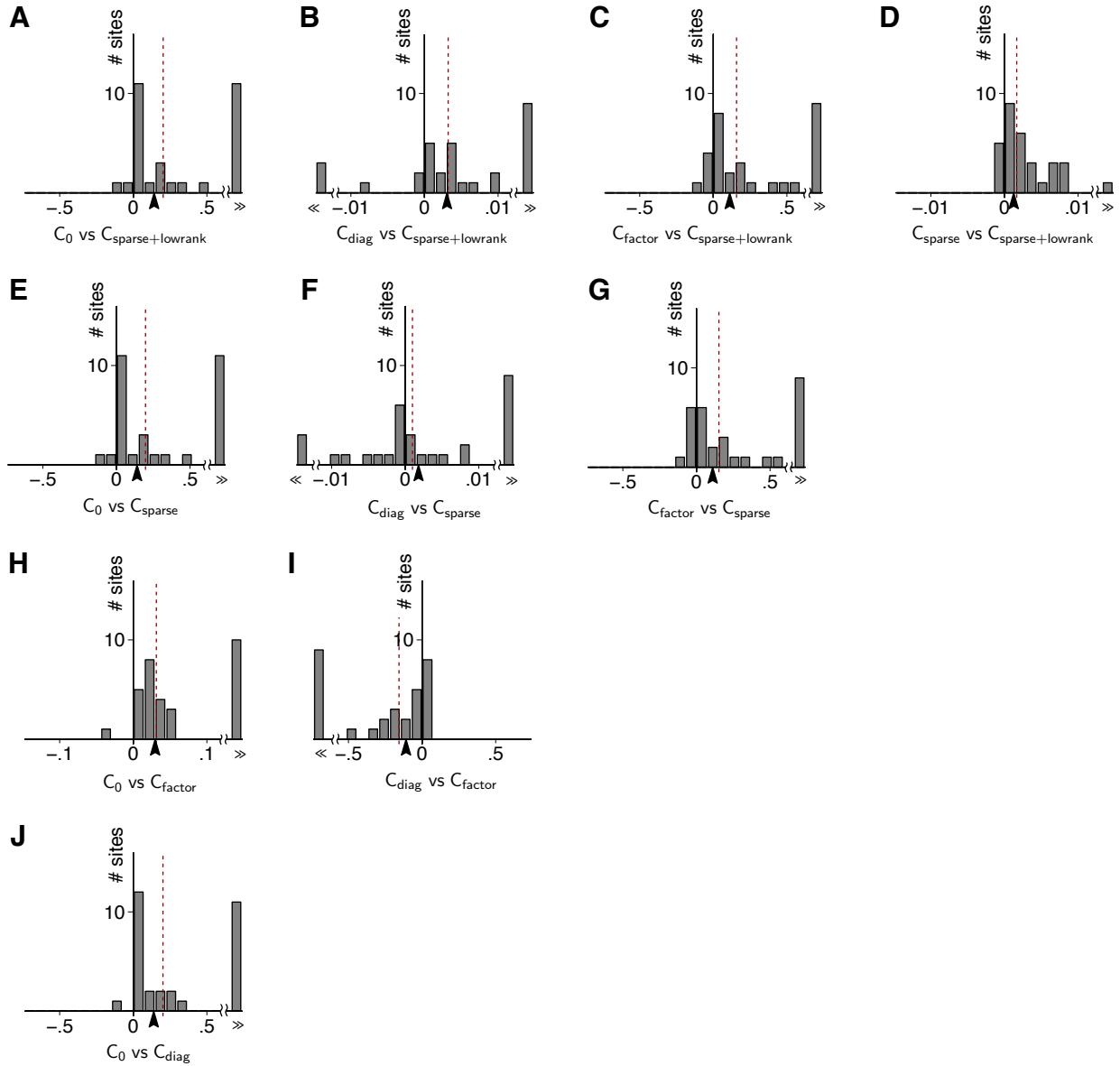
12. Gerstein G, Clark W (1964) Simultaneous studies of firing patterns in several neurons. *Science* 143: 1325–1327.
13. Smith MA, Kohn A (2008) Spatial and temporal scales of neuronal correlation in primary visual cortex. *J Neurosci* 28: 12591–12603.
14. Denman DJ, Contreras D (2013) The structure of pairwise correlation in mouse primary visual cortex reveals functional organization in the absence of an orientation map. *Cereb Cortex* .
15. Ko H, Hofer SB, Pichler B, Buchanan KA, Sjöström PJ, et al. (2011) Functional specificity of local synaptic connections in neocortical networks. *Nature* 473: 87–91.
16. Arieli A, Shoham D, Hildesheim R, 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: 2072–2093.
17. Chiu C, Weliky M (2002) Relationship of correlated spontaneous activity to functional ocular dominance columns in the developing visual cortex. *Neuron* 35: 1123–1134.
18. Kenet T, Bibitchkov D, Tsodyks M, Grinvald A, Arieli A (2003) Spontaneously emerging cortical representations of visual attributes. *Nature* 425: 954–956.
19. Cohen MR, Newsome WT (2008) Context-dependent changes in functional circuitry in visual area mt. *Neuron* 60: 162–173.
20. Cohen MR, Maunsell JH (2009) Attention improves performance primarily by reducing interneuronal correlations. *Nature neuroscience* 12: 1594–1600.
21. Rothschild G, Nelken I, Mizrahi A (2010) Functional organization and population dynamics in the mouse primary auditory cortex. *Nature neuroscience* 13: 353–360.
22. Smith MA, Sommer MA (2013) Spatial and temporal scales of neuronal correlation in visual area v4. *The Journal of Neuroscience* 33: 5422–5432.
23. Hofer SB, Ko H, Pichler B, Vogelstein J, Ros H, et al. (2011) Differential connectivity and response dynamics of excitatory and inhibitory neurons in visual cortex. *Nature neuroscience* 14: 1045–1052.
24. Hansen BJ, Chelaru MI, Dragoi V (2012) Correlated variability in laminar cortical circuits. *Neuron* 76: 590–602.
25. Smith MA, Jia X, Zandvakili A, Kohn A (2013) Laminar dependence of neuronal correlations in visual cortex. *Journal of neurophysiology* 109: 940–947.
26. Golshani P, Gonçalves JT, Khoshkhoo S, Mostany R, Smirnakis S, et al. (2009) Internally mediated developmental desynchronization of neocortical network activity. *J Neurosci* 29: 10890–9.
27. Gu Y, Liu S, Fetsch CR, Yang Y, Fok S, et al. (2011) Perceptual learning reduces interneuronal correlations in macaque visual cortex. *Neuron* 71: 750–761.
28. Greenberg DS, Houweling AR, Kerr JN (2008) Population imaging of ongoing neuronal activity in the visual cortex of awake rats. *Nature neuroscience* 11: 749–751.
29. Goard M, Dan Y (2009) Basal forebrain activation enhances cortical coding of natural scenes. *Nat Neurosci* 12: 1444–9.

30. Kohn A, Zandvakili A, Smith MA (2009) Correlations and brain states: from electrophysiology to functional imaging. *Curr Opin Neurobiol* 19: 434-8.
31. Renart A, de la Rocha J, Bartho P, Hollender L, Parga N, et al. (2010) The asynchronous state in cortical circuits. *Science* 327: 587-90.
32. Perkel DH, Gerstein GL, Moore GP (1967) Neuronal spike trains and stochastic point processes: II. simultaneous spike trains. *Biophysical journal* 7: 419–440.
33. Moore GP, Segundo JP, Perkel DH, Levitan H (1970) Statistical signs of synaptic interaction in neurons. *Biophysical Journal* 10: 876–900.
34. Shadlen MN, Newsome WT (1998) The variable discharge of cortical neurons: implications for connectivity, computation, and information coding. *J Neurosci* 18: 3870-96.
35. Salinas E, Sejnowski TJ (2001) Correlated neuronal activity and the flow of neural information. *Nature Reviews Neuroscience* 2: 539–550.
36. Ostojic S, Brunel N, Hakim V (2009) How connectivity, background activity, and synaptic properties shape the cross-correlation between spike trains. *The Journal of Neuroscience* 29: 10234–10253.
37. Rosenbaum R, Josić K (2011) Mechanisms that modulate the transfer of spiking correlations. *Neural computation* 23: 1261–1305.
38. 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: 32.
39. Peng J, Wang P, Zhou N, Zhu J (2009) Partial correlation estimation by joint sparse regression models. *Journal of the American Statistical Association* 104.
40. Varoquaux G, Gramfort A, Poline JB, Thirion B (2012) Markov models for fmri correlation structure: is brain functional connectivity small world, or decomposable into networks? *Journal of Physiology-Paris* 106: 212–221.
41. Ryali S, Chen T, Supekar K, Menon V (2012) Estimation of functional connectivity in fmri data using stability selection-based sparse partial correlation with elastic net penalty. *Neuroimage* 59: 3852–3861.
42. Ledoit O, Wolf M (2004) A well-conditioned estimator for large-dimensional covariance matrices. *Journal of multivariate analysis* 88: 365–411.
43. Bickel PJ, Li B, Tsybakov AB, van de Geer SA, Yu B, et al. (2006) Regularization in statistics. *Test* 15: 271–344.
44. Ledoit O, Wolf M (2003) Improved estimation of the covariance matrix of stock returns with an application to portfolio selection. *Journal of Empirical Finance* 10: 603–621.
45. Chandrasekaran V, Parrilo PA, Willsky AS (2010) Latent variable graphical model selection via convex optimization. In: *Communication, Control, and Computing (Allerton), 2010 48th Annual Allerton Conference on*. IEEE, pp. 1610–1613.
46. Ma S, Xue L, Zou H (2013) Alternating direction methods for latent variable gaussian graphical model selection. *Neural computation* : 1–27.

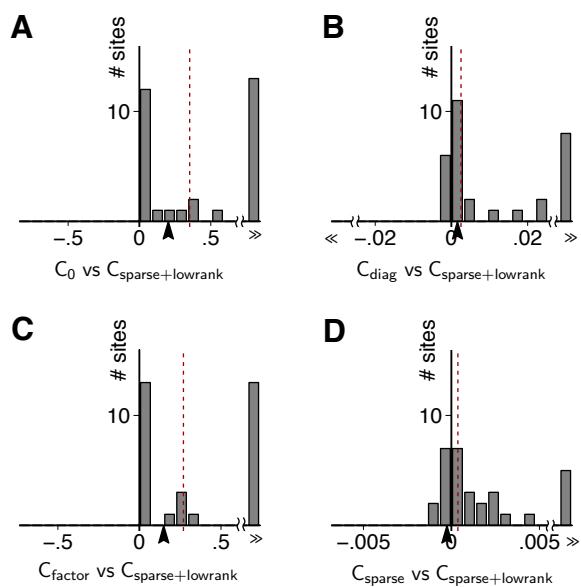
47. James W, Stein C (1961) Estimation with quadratic loss. In: Proceedings of the fourth Berkeley symposium on mathematical statistics and probability. volume 1, pp. 361–379.
48. Fan J, Fan Y, Lv J (2008) High dimensional covariance matrix estimation using a factor model. *Journal of Econometrics* 147: 186–197.
49. Reddy GD, Saggau P (2005) Fast three-dimensional laser scanning scheme using acousto-optic deflectors. *J Biomed Opt* 10: 064038.
50. Katona G, Szalay G, Maák P, Kaszás A, Veress M, et al. (2012) Fast two-photon in vivo imaging with three-dimensional random-access scanning in large tissue volumes. *Nat Methods* .
51. Cotton RJ, Froudarakis E, Storer P, Saggau P, Tolias AS (2013) Three-dimensional mapping of microcircuit correlation structure. *Frontiers in Neural Circuits* 7: 151.
52. Vogelstein JT, Packer AM, Machado TA, Sippy T, Babadi B, et al. (2010) Fast nonnegative deconvolution for spike train inference from population calcium imaging. *Journal of neurophysiology* 104: 3691–3704.
53. Malmersjö S, Rebellato P, Smedler E, Planert H, Kanatani S, et al. (2013) Neural progenitors organize in small-world networks to promote cell proliferation. *Proceedings of the National Academy of Sciences* 110: E1524–E1532.
54. Alonso JM, Martinez LM (1998) Functional connectivity between simple cells and complex cells in cat striate cortex. *Nature neuroscience* 1: 395–403.
55. Aertsen A, Gerstein G, Habib M, Palm G (1989) Dynamics of neuronal firing correlation: modulation of “effective connectivity”. *Journal of neurophysiology* 61: 900–917.
56. Pillow JW, Shlens J, Paninski L, Sher A, Litke AM, et al. (2008) Spatio-temporal correlations and visual signalling in a complete neuronal population. *Nature* 454: 995–999.
57. Schneidman E, Berry MJ 2nd, Segev R, Bialek W (2006) Weak pairwise correlations imply strongly correlated network states in a neural population. *Nature* 440: 1007-12.
58. Tkacik G, Schneidman E, Berry I, Michael J, Bialek W (2006) Ising models for networks of real neurons. arXiv preprint q-bio/0611072 .
59. Yu S, Huang D, Singer W, Nikolic D (2008) A small world of neuronal synchrony. *Cereb Cortex* 18: 2891-901.
60. Tang A, Jackson D, Hobbs J, Chen W, Smith JL, et al. (2008) A maximum entropy model applied to spatial and temporal correlations from cortical networks in vitro. *J Neurosci* 28: 505-18.
61. Shlens J, Field GD, Gauthier JL, Greschner M, Sher A, et al. (2009) The structure of large-scale synchronized firing in primate retina. *J Neurosci* 29: 5022-31.
62. Feldt S, Bonifazi P, Cossart R (2011) Dissecting functional connectivity of neuronal microcircuits: experimental and theoretical insights. *Trends in neurosciences* 34: 225–236.
63. Grewe BF, Langer D, Kasper H, Kampa BM, Helmchen F (2010) High-speed in vivo calcium imaging reveals neuronal network activity with near-millisecond precision. *Nat Meth* 7: 399–405.
64. Leung LC, Wang GX, Mourrain P (2013) Imaging zebrafish neural circuitry from whole brain to synapse. *Frontiers in neural circuits* 7.

65. Ahrens MB, Orger MB, Robson DN, Li JM, Keller PJ (2013) Whole-brain functional imaging at cellular resolution using light-sheet microscopy. *Nature methods* 10: 413–420.
66. Perin R, Berger TK, Markram H (2011) A synaptic organizing principle for cortical neuronal groups. *Proc Natl Acad Sci U S A* 108: 5419–24.
67. Gerstein GL, Bedenbaugh P, Aertsen AM (1989) Neuronal assemblies. *Biomedical Engineering, IEEE Transactions on* 36: 4–14.
68. Chapin JK, Nicolelis MA (1999) Principal component analysis of neuronal ensemble activity reveals multidimensional somatosensory representations. *Journal of neuroscience methods* 94: 121–140.
69. Peyrache A, Benchenane K, Khamassi M, Wiener SI, Battaglia FP (2010) Principal component analysis of ensemble recordings reveals cell assemblies at high temporal resolution. *Journal of computational neuroscience* 29: 309–325.
70. Ch'Ng YH, Reid RC (2010) Cellular imaging of visual cortex reveals the spatial and functional organization of spontaneous activity. *Frontiers in integrative neuroscience* 4.
71. Lopes-dos Santos V, Conde-Ocazinez S, Nicolelis MA, Ribeiro ST, Tort AB (2011) Neuronal assembly detection and cell membership specification by principal component analysis. *PloS one* 6: e20996.
72. Lopes-dos Santos V, Ribeiro S, Tort AB (2013) Detecting cell assemblies in large neuronal populations. *Journal of Neuroscience Methods* 220: 149 – 166.
73. Okun M, Yger P, Marguet SL, Gerard-Mercier F, Benucci A, et al. (2012) Population rate dynamics and multineuron firing patterns in sensory cortex. *J Neurosci* 32: 17108–19.
74. Köster U, Sohl-Dickstein J, Gray CM, Olshausen BA (2013) Higher order correlations within cortical layers dominate functional connectivity in microcolumns. *arXiv preprint arXiv:13010050*.
75. Ganmor E, Segev R, 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: 9679–84.
76. Tkačik G, Marre O, Amodei D, Schneidman E, Bialek W, et al. (2013) Searching for collective behavior in a network of real neurons. *arXiv preprint arXiv:13063061*.
77. Roudi Y, Nirenberg S, 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: e1000380.
78. Alivisatos AP, Chun M, Church GM, Deisseroth K, Donoghue JP, et al. (2013) The brain activity map. *Science* 339: 1284–1285.
79. Efron B, Morris CN (1977) Stein's paradox in statistics. WH Freeman.
80. Schäfer J, Opgen-Rhein R, Strimmer K (2010) *corpcor: Efficient estimation of covariance and (partial) correlation*. R package version 15.7.
81. Meinshausen N, Bühlmann P (2006) High-dimensional graphs and variable selection with the lasso. *The Annals of Statistics* 34: 1436–1462.
82. Friedman J, Hastie T, Tibshirani R (2008) Sparse inverse covariance estimation with the graphical lasso. *Biostatistics* 9: 432–441.

## Supporting Information



**Figure S 1.** All-to-all performance comparisons of the sample covariance matrix and the four regularized estimators with respect to multivariate normal cross-validation loss.



**Figure S 2.** The sparse+latent estimator outperforms the other estimators under the quadratic loss (Eq. 14) just as it did under the normal loss (Eq. 9) used in Fig. 3.