

Improved estimation of neural correlations suggests detailed interactions in visual cortex

Dimitri Yatsenko,^{1,*}, Krešimir Josić², Alexander S. Ecker^{1,3,4}, Emmanouil Froudarakis¹, R. James Cotton¹, Andreas S. Tolias^{1,5}

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@cnsbcm.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 functional organization of neural circuits. However, estimating and interpreting large correlation matrices obtained from finite recordings is challenging. Estimation can be improved by regularization: imposing some kind of structure on the estimate. The amount of improvement depends on how well the imposed structure accommodates the dominant dependencies in the data. Therefore, the selection of the most efficient estimator of neural correlation matrices is an empirical question that informs about the types of dominant dependencies governing the system.

In this study, we sought most statistically efficient estimators of neural correlations 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 μm wide and 100 μm deep (150–350 cells) in mouse visual cortex. We hypothesized that in these dense recordings, the correlation matrix would be most efficiently represented by a sparse network of partial correlations between pairs of observed neurons combined with common latent factors representing unobserved inputs and global network fluctuations. Indeed, in cross-validation tests, the covariance matrix estimator based on such approximations outperformed other regularized estimators. Mirroring previously described patterns of synaptic connectivity, the density of positive partial correlations decreased rapidly with both physical distances and preferred orientation differences whereas negative partial correlations were less selective. The inferred functional connectivity obtained using these methods can generate hypotheses about synaptic architecture that can guide future experimental work.[1]

Author Summary

Correlations in the activity of neurons have proven useful as descriptors of the functional organization of neural circuits with implications for stimulus coding and circuit architecture. Estimation of correlation matrices can be improved by imposing some structure on the estimate. Greatest improvements are attained when the imposed structure efficiently captures real dependencies in the data. Using fast 3D two-photon imaging of calcium signals, we recorded the activity of large and dense groups of cells in mouse visual cortex and evaluated the performance of correlation matrix estimators that imposed different kinds of structure. The correlation structure of the estimator that led to greatest improvement comprised a sparse network of partial correlations between pairs of neurons combined with partial correlations with several common latent factors. Both components were necessary for efficient estimation. Since it leads to

better estimates, we proposed that this sparse+latent structure provides a better picture of the functional connectivity than that obtained from the usual sample correlation matrix in densely sampled neural recordings. As an application of this approach, we analyzed how the inferred connectivity related to distances between cells and differences in their preferred orientations and found basic agreement with previous studies of synaptic connectivity.

Introduction

Neural correlations Pearson correlations between the spiking activity of pairs of neurons, or simply *neural correlations*, are among the most familiar descriptive statistics of neural population activity [1–5]. For example, *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, 6–10]. In addition, noise correlations and correlations in spontaneous activity have been hypothesized to reflect key features of functional connectivity in neural circuits[2]. 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 [11, 12], their synaptic connectivity [13], stimulus response similarity [?, 3, 4, 13–18], cortical layer specificity [19, 20], progressive changes in development and in learning [21, 22], changes due to sensory stimulation and global brain states [5, 23–25], and others.

However, neural correlations do not submit to ready or unambiguous mechanistic interpretation. Theoretical studies and simulations have shown that neural correlations on various temporal scales may arise from combinations of multiple mechanisms including direct synaptic interactions, common inputs or correlated inputs, chains of multiple synaptic connections, oscillations, top-down modulation, and background network fluctuations [26–31].

Additional information in neural correlation matrices Yet, a complete [4]correlation matrix provides more information than the equivalent number of isolated pairwise correlations. In early studies, the effects of correlations were extrapolated from isolated pairs to entire populations through simulation and theoretical analysis [2, 28]. In contrast, modern multineuronal recordings from increasingly large and dense populations of highly interconnected neurons allow estimation of various low-dimensional representations of the entire correlation matrix, which can enable inquiry into the nature of underlying physiological mechanisms and computational principles.

For example, one form of low-dimensional representation of the correlation matrix is its low-rank approximation. Such approximations are particularly suitable for capturing shared fluctuations across the recorded population.[5] Low-rank components can be extracted using principal component or factor analysis. The resulting structure of neural correlations has been analyzed in a number of studies with particular attention paid to their temporal dynamics [?]. [7]

Alternatively, the correlation matrix can be expressed through the partial pairwise correlations. The partial pairwise correlation is the linear correlation remaining[8] after accounting for correlations with all the other neurons. Partial correlations carry particular significance when the relationship between the variables is approximately linear. In this case, small pairwise partial correlations indicate that the two variables are nearly conditionally independent[9].

Conditional independence between a pair of variables suggests a lack of direct interaction between the underlying processes. Therefore, networks of partial correlations (sometimes called *association networks*) can be used to infer interactions between the components of a complex system. This approach has been used to uncover the structure of gene interaction networks [32, 33] and the functional connectivity between brain regions from fMRI signals [34, 35]. In *elliptical distributions* [10](those of the form $x \sim \frac{1}{Z}g((x - \mu)^T \Sigma^{-1}(x - \mu))$, where g is an arbitrary function, Z is the scalar normalizer, Σ is the covariance matrix) and the multivariate normal distribution in particular, interactions are completely defined by linear effects and partial correlations express conditional dependencies. A multivariate normal distribution

specified by a weighted graph[11] of non-zero partial correlations is known as a Gaussian Graphical Model or Gauss-Markov Random Field [36]. With departure from linearity, the correspondence between conditional dependence and partial correlations diminishes and can break down [37]. For example, the coupling terms in pairwise Ising models have no direct relationship to the partial correlations [38,39].[12]

Estimation of neural correlation matrices Here, we pursued two related aims: (a) the improved estimation of neural correlation matrices and (b) discovery of the low-dimensional structure of correlations in recordings of multineuronal activity. An accurate characterization of the patterns of correlations helps in interpreting them, and to better relate them to the architecture and dynamics of the underlying neuronal networks.

The true correlation matrix is the normalized version of the true covariance matrix defined as

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

where $\mathbb{E} [\cdot]$ denotes expectation; x is the $p \times 1$ vector of real-valued instantaneous firing rates in bins of duration Δt ; and μ is the vector of mean firing rates. In the case of noise correlations, μ depends on the stimulus condition. The usual estimator [13] of the covariance matrix is the *sample covariance matrix* C_0 computed 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 ν is the number of degrees of freedom per neuron in the sample ($\nu = n - 1$ if observations are independent). In this study, we estimate the true mean, μ , using the sample mean \bar{x} , but seek a better estimate of the covariance matrix than C_0 .

Given a covariance matrix estimate C (not necessarily the sample covariance C_0), the correlation matrix R is calculated by normalizing C by its diagonal (variance estimates):

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

Similarly, the matrix of partial correlations P is computed as the negative normalized version of the *precision matrix* (inverse covariance matrix) C^{-1} :

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

Here \circ denotes entrywise matrix product (Hadamard product) and I is the $p \times p$ identity matrix. Clearly, off-diagonal zeros in the precision matrix correspond to zero partial correlations.

Estimation of large covariance matrices 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 shared activity (*i.e.* overestimation of large eigenvalues) [40], and poorly conditioned estimates of the partial correlations [32].

Estimation can be improved through *regularization*: the deliberate biasing of the estimate toward a low-dimensional approximation [32, 41]. The sample estimate C_0 is unbiased ($\mathbb{E}[C_0] = \Sigma$) but, on average, falls far from Σ due to its sensitivity to sampling noise. Low-dimensional estimates are typically less susceptible to sampling noise but can introduce biases. Regularization strikes a favorable balance between bias and variability, and can produce *some* improvement even with an arbitrary target estimate. [14] Yet, when the target estimate captures important features of the true covariance matrix with few terms, a regularized estimate introduces minimal bias and outperforms other estimators.

Illustration of estimation of a neural correlation matrix [15] To illustrate the challenges in estimating the correlation matrix from a finite sample, we consider a regularization scheme based on *covariance selection* (Figure 1) [42]. Here the estimate is produced by fitting only an optimal subset of the coefficients of the precision matrix while setting the rest to zero. We estimated the covariance matrix of somatic calcium signals of a group of 298 neurons in mouse visual cortex using both the sample covariance matrix (Fig. 1 A, B) and covariance selection (Fig. 1 C). Due to the low-pass filtration effect of calcium dye kinetics, correlations in unprocessed calcium signals are higher than typical firing rate correlations on shorter temporal scales. The close similarity of the two estimates of the correlation matrix (Fig. 1 B and C) belies the utterly different partial correlation structure (Fig. 1 D and E): The regularized estimate is produced from the same data by setting to zero 31,501 (71.2%) of the off-diagonal coefficients of the precision matrix and fitting the remaining 12,752 coefficients. Estimation by covariance selection is attractive for several reasons: First, with fewer free parameters it is less susceptible to sampling noise, yielding estimates that are, on average, closer to the truth. Additionally, if we assume that partial correlations do indeed reflect elements of the functional connectivity in the circuit, the second estimate also appears more plausible: in neural circuits, each neuron only interacts with a fraction of the entire population.

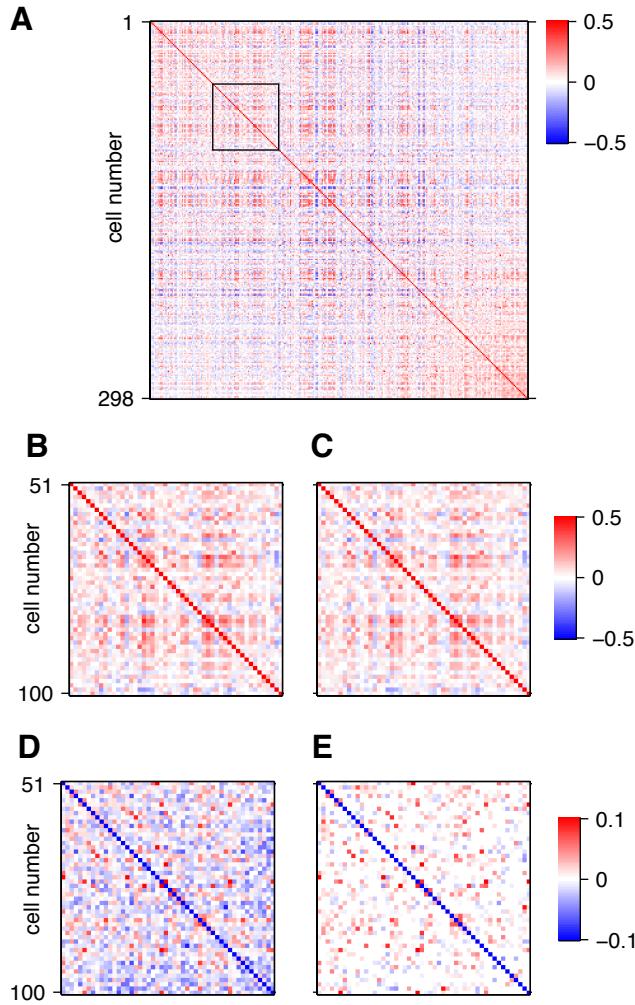


Figure 1. Illustration of regularized estimation of partial correlations. **A.** The sample correlation matrix of unprocessed somatic calcium signals from a population of cells in mouse visual cortex. The outlined square fragment is magnified in **B**. **C.** The same fragment of another estimate of the correlation matrix regularized to yield sparse partial correlations. Corresponding fragments of partial correlations matrices of the unregularized and regularized estimated are shown in **D** and **E**, respectively.

Model selection Covariance selection is but one of many conceivable low-dimensional approximations of the covariance matrix. Indeed any probabilistic model of the joint population activity regularized toward a small number of parameters automatically defines a low-dimensional approximation of the covariance matrix. Models that accurately reflect the dominant regularities in the activity of the circuit, its so-called ‘functional connectivity’, tend to produce estimates that are nearest to the true value of the covariance matrix and are said to be more *efficient* or to *dominate* other estimators. An estimator’s superior efficiency can be ascertained by cross-validation. Because different types of neural circuits have different types of regularities in their activity, different probabilistic models and covariance matrix estimators may dominate in each case. The estimator that is shown to dominate all others for a specific type of neural circuit may provide [16] the best descriptor of the low-dimensional representation of the correlation structure of the system of interest. Its structure may then be analyzed and related to the circuit’s anatomical organization.

Summary of findings In this study, we compared four regularized covariance matrix estimators biased toward different respective low-dimensional correlation structures: ‘shrinkage toward diagonal’ or C_{diag} , ‘shrinkage toward a multifactor model’ or C_{factor} , ‘sparse partial correlations’ or C_{sparse} , and ‘sparse partial correlations with latent units’ or $C_{\text{sparse+latent}}$. First, we demonstrated that, in simulations with known true low-dimensional correlation structures, regularized estimators with matching types of low-dimensional structure generally dominated over the other estimators. We then used cross-validation to determine the most efficient estimator for the population activity of dense groups of neurons in mouse primary visual cortex. We found that $C_{\text{sparse+latent}}$ consistently dominated the other estimators. Estimate produced by $C_{\text{sparse+latent}}$ revealed networks of interactions that differed substantially from networks of strongest correlations and depended strongly on the physical distance separating pairs of cells and on the differences in their preferred orientations. [17]

Results

Covariance estimation We considered four regularized estimators based on distinct families of low-dimensional target estimates: ‘independent’, ‘latent factors’, ‘sparse interactions’, and ‘sparse+latent’ [18] (Fig. 2 row 1).

For the first regularized estimator C_{diag} , the target estimate is the diagonal matrix D whose diagonal contains variance estimates. Regularization is achieved by linear *shrinkage* of the unbiased estimate C_0 toward D controlled by the scalar *shrinkage intensity* parameter $\lambda \in [0, 1]$:

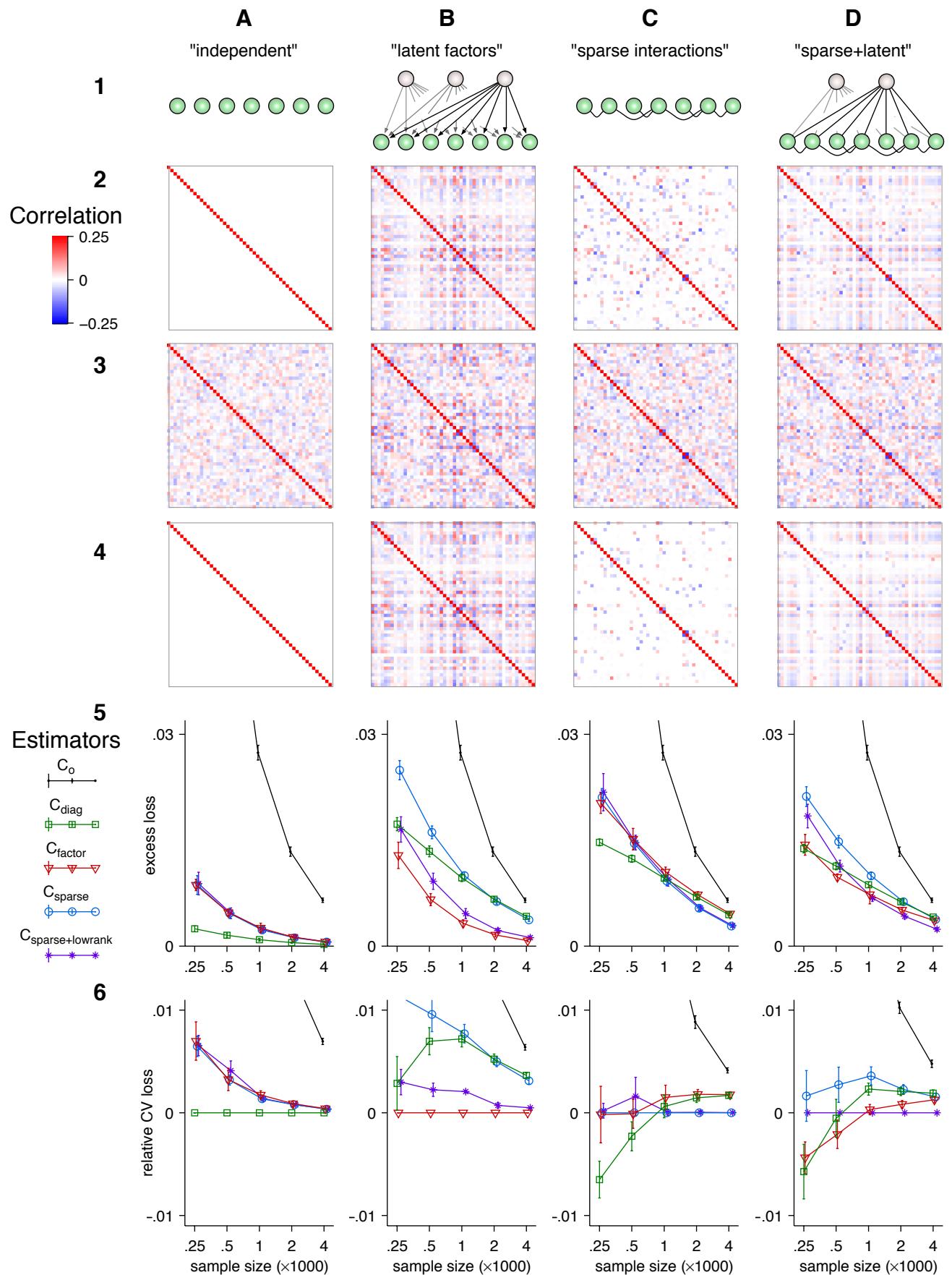
$$C_{\text{diag}} = (1 - \lambda)C_0 + \lambda D \quad (5)$$

The diagonal target has structure corresponding to the absence of linear associations between the activity of observed neurons (Fig. 2 A). If this structure accurately describes recorded data, then strong shrinkage toward D adds little bias while strongly reducing the variability of the estimate. Shrinkage allows for partial commitment to the low-dimensional representation.

The target of the second regularized estimator C_{factor} , is the factor model $LL^T + \Psi$, where L is the $p \times d$ matrix of *factor loadings*, where d is the number of factors, and Ψ is the diagonal matrix of independent variances. The estimator,

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

has two hyperparameters: the number of factors d and shrinkage intensity λ . The target estimate $LL^T + \Psi$ has the structure that arises when correlated fluctuations in population activity are linearly driven by a small number [19] of latent factors that affect many cells while direct interactions between cells are insignificant (Fig. 2 B).



The third estimator C_{sparse} is based on the assumption that all correlations result from direct linear associations between a sparse set of pairs of observed cells (Fig. 2 C). This assumption is enforced by setting to zero an optimal[20] subset of the off-diagonal elements of the precision matrix. Therefore, C_{sparse} is biased towards correlation structures that arise in networks of linearly interacting pairs of neurons (Fig. 2 C). The estimator takes the form

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

where S is a sparse matrix with a large fraction of zeros off the diagonal. The estimate has one hyperparameter that determines the sparsity (fraction of off-diagonal zeros) of S .

The fourth estimator $C_{\text{sparse+latent}}$, combines the assumptions of the previous two. The underlying structure arises from sparse interactions between the recorded neurons and latent factors driving their activity (Fig. 2 D). The estimator has the form [43, 44]:

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

where, as above, S is a sparse matrix and L is a $d \times p$ matrix of factor loadings. The estimator has two hyperparameters: the number of latent units d and the sparsity of S .

Simulation To illustrate the performance of the four regularized estimators, we constructed four model populations of $p = 50$ neurons each with each matching the correlation structures of a different target estimate (Fig. 2 Row 2). Sample correlation matrices calculated from samples of size $n = 1000$ contain substantial sampling noise (Fig. 2 Row 3) that makes the correlation structure less obvious.

To evaluate the performance of a covariance matrix estimator C , we define a real-valued *loss function* $\mathcal{L}(C, \Sigma)$ that attains a minimum when $C = \Sigma$. The loss function quantifies the error of the estimate, *i.e.* the deviation of the estimate C from truth Σ .

In this study, we adopted *negative normal log-likelihood loss* defined as

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

Figure 2 (preceding page). Estimators whose low-dimensional regularization targets can represent the structure of the true covariance matrix outperform other estimators. **Row 1.** Graphical representations of four types of low-dimensional structures of interactions between observed neurons (green spheres) and latent units (light-shaded spheres). In the ‘independent model’ (**A**), observed neurons exert no linear effects on one another neither directly nor through interactions with common latent units. In ‘latent factors’ (**B**), the correlated activity of observed cells is driven by several latent units. In ‘sparse interactions’ (**C**), the correlation matrix is defined by a set of linear interactions between observed neurons. In ‘sparse+latent’ (**D**), correlations arise through direct linear interactions in a sparse subset of observed neurons and through interactions with common latent units. **Row 2.** Examples of 50×50 correlation matrices corresponding to each type of low-dimensional structure. The factor model (**B**) has three latent units. The partial correlation matrix of the sparse model (**C**) is 73% sparse. The ‘sparse+latent’ matrix has one latent unit and its direct interactions are 78% sparse. **Row 3.** Examples of sample correlation matrices calculated from samples of 1000 observations taken from simulated random processes with corresponding correlation matrices from row 2. **Row 4.** Mean *excess loss* (Eq. 11) attained by each of the five estimators as a function of sample size. The error bars indicate the standard error of the mean based on 30 samples. **Row 5.** Mean *cross-validation loss* (Eq. 13) of covariance estimators with respect to the matching estimator. The values are relative to the validation loss of the estimator that matches the low-dimensional structure of the true covariance matrix. Error bars indicate the standard error of the mean based on 30 samples.

The loss function $\mathcal{L}(C, \Sigma)$ is related to the multivariate log-likelihood function $L(\Sigma|C_0)$ as

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

We normalized the loss function by $\frac{1}{p}$ to make its value comparable across different population sizes.

The choice of 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 the Frobenius norm of the difference $\|C - \Sigma\|_F$ [32, 40], Stein's entropy loss, or quadratic loss [45, 46].

We define the *excess loss* as

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

which assumes zero at its minimum.

In Figure 2, row 4 shows the mean excess losses and their standard errors calculated from 30 samples of sizes $n=250, 500, 1000, 2000$, and 4000 for each estimator and each ground truth. For each estimator, hyperparameters were estimated by nested cross-validation (See Methods for more details.)

As expected, estimators with whose structure matched that of the true model typically outperformed the other estimators. There were two exceptions: First, for small sample sizes, there was insufficient data to reveal the true correlation structure. In this case estimates with simpler targets often outperformed the estimator based on the correct model. Second, the sparse+latent estimator often performed equally well to the sparse estimator even when the data did not have any latent units. In these cases, it inferred the right number of latent units: zero. In larger models, the falsely identified latent factors are more common.

When the ground truth, Σ , is not accessible, loss can be estimated solely from the data through *validation*. In validation, an additional independent *testing sample* is used to compute a sample covariance estimate C'_0 . Then *validation loss* $\mathcal{L}(C, C'_0)$ can be used as an unbiased estimate¹ of loss $\mathcal{L}(C, \Sigma)$. Thus, estimators resulting in consistently lower validation loss can be inferred to produce estimates that are closer to truth than estimators with higher validation loss.

With empirical data, acquiring a separate testing sample is not practical. Instead, K -fold cross-validation is 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 a training dataset. The validation losses from each of such ‘folds’ are averaged to produce *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 the present formulation, ℓ_C is known as the cross-validated Gaussian log-likelihood (up to a constant offset and multiplier).

¹Validation loss $\mathcal{L}(C, C'_0)$ is an unbiased estimate of loss $\mathcal{L}(C, \Sigma)$ when $\mathcal{L}(\cdot, \cdot)$ is *additive* in its second argument so that

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

Then the absence of bias can be shown:

$$\mathbb{E}_C \left[\mathbb{E}_{C'_0} [\mathcal{L}(C, C'_0)] \right] = \mathbb{E}_C [\mathcal{L}(C, \mathbb{E}_{C_0} [C'_0])] = \mathbb{E}_C [\mathcal{L}(C, \Sigma)]$$

The last result expression is known as the *estimator risk* which serv

The property of additivity does not hold for other popular loss functions such as Stein's entropy loss or various quadratic losses, for example, which prevents their substitution with corresponding validation losses.

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

$$\ell_{C,C_{\text{ref}}} = \frac{1}{K} \sum_{k=1}^K \left[\mathcal{L}\left(C^{\{\setminus k\}}, C_0^{\{k\}}\right) - \mathcal{L}\left(C_{\text{ref}}^{\{\setminus k\}}, C_0^{\{k\}}\right) \right] \quad (13)$$

In simulation, CV-loss accurately reproduced the differences between the estimators' excess losses, although with greater variability (Figure 2, Row 5). For each kind of ground truth (Row 2), relative CV-losses were computed with respect to the estimator whose regularization target matched the structure of the respective ground truth. Just as with excess loss in Row 4, the means and their standard errors were computed from 30 samples taken for each ground truth and for each sample size.[21]

These simulations demonstrate that with sufficiently large sample sizes the most efficient among several regularized estimators could be used to infer the most likely type of low-dimensional structure present in the data.

Covariance estimation in neural data We recorded calcium activity in dense populations of neurons in the supragranular layers in primary visual cortex of anesthetized mice using fast random-access 3D scanning two-photon microscopy [47–49]. We presented 300 repetitions of full-field drifting gratings with two directions of motion or 150 repetitions with five directions (Fig. 3 A) on one side of the visual field of anesthetized mice. Groups of cells loaded with a calcium-sensitive fluorescent dye were imaged and localized in 3D space (Fig. 3 B and E in the visual cortex on the contralateral side to the stimulus. Using acousto-optic deflectors (AOD) to steer the laser in 3D, we recorded the somatic calcium activity from the located cells with concurrent motion detection [49]. [22] 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). After downsampling the signal to 20 Hz, relative firing rates were inferred using sparse nonnegative deconvolution [50] (Fig. 3 C and D). Only cells that produced detectable calcium activity were included in the analysis[23] (See Methods).[24] The average stimulus response was subtracted from each trial; the remaining signals were further downsampled into 150 ms bins to compute the noise correlation matrix (Fig. ?? F and G). The durations of recordings dedicated to estimating the noise correlations ranged between 15–20 minutes. Besides the high-repetition stimulus protocol used to estimate noise correlations, an orientation tuning stimulus protocol was used to map the orientation tuning of each cell (Fig. 3 E). Overall, 31 imaged sites from 24 animals were included in the study.

In these highly localized populations both direct interactions between cells and common diffuse inputs likely contribute to overall population co-variability. We therefore hypothesized that regularized estimates capturing a sparse partial correlation structure, and/or interactions with latent units would be the most efficient. At the same time, average correlations were relatively small (Fig. 3 G), giving a potential advantage to estimates biased toward independence (*e.g.* C_{diag}). Finally, correlation structure could possibly be best explained by common fluctuations across the entire population, to the advantage of estimates biased toward low-rank correlation structure (*e.g.* C_{factor} , $C_{\text{factor+lowrank}}$). Thus it was *a priori* unclear which estimator would perform best.

We found that estimator $C_{\text{sparse+latent}}$ was the most efficient (Fig. 4). The median relative CV-validation loss (Eq. 13) of each of the estimates C_0 , C_{diag} , C_{factor} , and C_{sparse} with respect to $C_{\text{sparse+latent}}$ was significantly above zero ($p < 0.001$, Wilcoxon signed rank test, for each).

Correlation structure and circuit architecture Having demonstrated that $C_{\text{sparse+latent}}$ dominated the other estimators, we examined the low-dimensional structure revealed by this estimator for individual imaged sites. By design, the sparse+latent estimator finds the optimal[25] balance between a sparse network of partial correlations and shared common latent units. If this estimate dominates, we can hypothesize that it is better reveals underlying physiological interactions than the usual correlations[26].

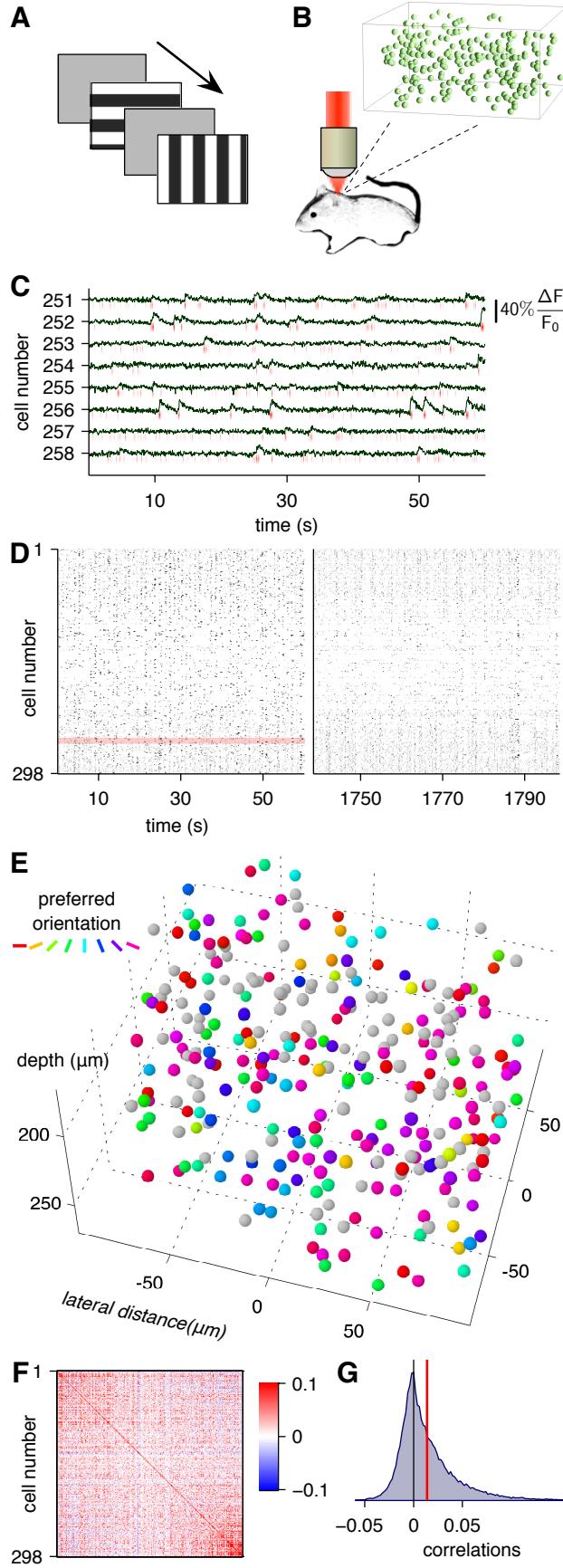


Figure 3. Acquisition of neural signals for the estimation of noise correlations. Visual stimuli comprising full-field drifting gratings interleaved with blank screens (A) were presented to anesthetized mice while two-photon recordings of somatic calcium signals were collected using fast 3D random-access microscopy (B). The visual stimulus included an initial period with 16 directions of motion for orientation tuning followed by a longer (15–20 min) period of stimulation with only 2 or 5 directions of motion for the computation of the noise correlation matrix. 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.

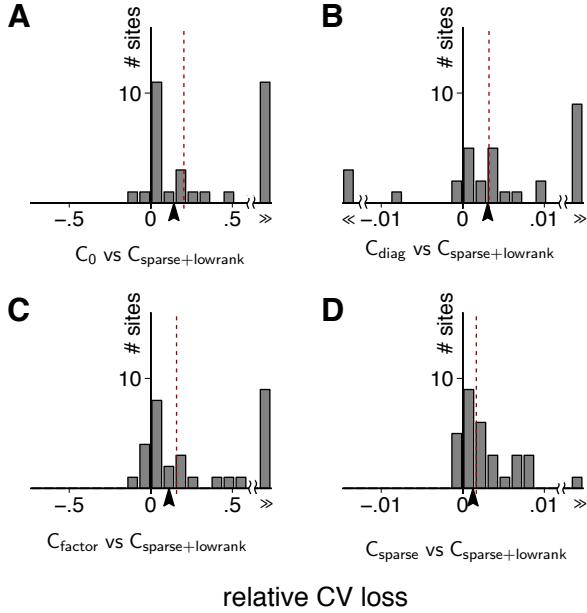


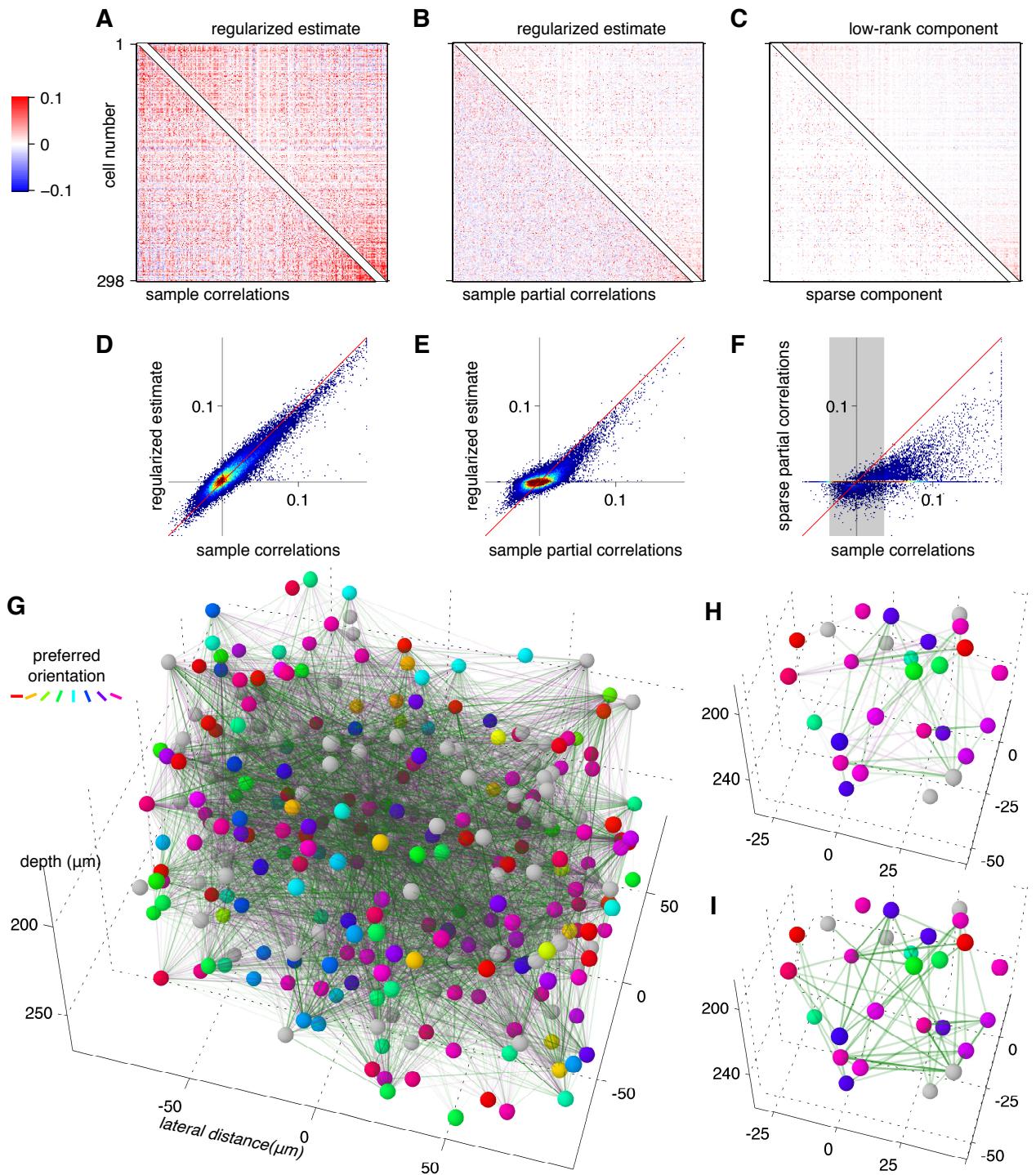
Figure 4. The sparse+latent estimator $C_{\text{sparse}+\text{latent}}$ outperforms the other estimators on neural data. **A–D.** Histograms of average cross-validation loss differences of the respective estimators C_0 , C_{diag} , C_{factor} , and C_{sparse} from $C_{\text{sparse}+\text{latent}}$. The histograms are based on 31 imaged sites in 24 mice. All medians (red dashed lines) were significantly greater than zero, indicating the dominance of $C_{\text{sparse}+\text{latent}}$ over the other estimators. The green arrows indicate the results for the site shown in Fig. 3 and Fig. 5

[27] In particular, the sparse component of the partial correlation matrix suggests direct interactions between pairs of neurons, whereas the low-rank component suggest common fluctuations such as those caused by common inputs or other collective synchronized activity.

We examined the structure of the sparse+latent estimate at individual sites (an example site is depicted in Fig. 3). At these sites the regularized estimate of the correlation matrix appeared very similar to the unregularized sample correlation matrix (Fig. 5 A and D). However, the corresponding partial correlation matrices differed dramatically (Fig. 5 B and E). The partial correlation was decomposed into sparse and low-rank components (Fig. 5 C). Although correlations were mostly positive, the sparse partial correlations (or ‘interactions’), had a much larger fraction of negative values than sample correlations. [28] The sparse component had 82.1% sparsity (or, conversely, 17.9% connectivity), which corresponded to the average node degree (interactions per cell) of 52.5 (Fig. 5 G). The low-rank component had rank 17. [29]

In previous studies of the structure of correlations, a sparse network was produced by thresholding the correlations coefficients at a level deemed significant. A network was formed by connecting neurons whose total correlations exceed this threshold [21, 51]. There was fairly little overlap between the network of interactions revealed by such thresholding and those revealed by the sparse+latent estimator (Fig. 5 F). When thresholded to the same sparsity (82.1%), only 42% of cell pairs connected in one network were connected in the other while the average magnitude of such correlations was much lower in the case of the regularized estimator (Fig. 5 F, H, and I). In particular, many low sample correlations translated into negative interactions in the regularized estimate. Indeed, the absence of correlations between pairs of cells that both correlate similarly to several of their neighbors should be considered as significant as a high correlation coefficient. [30] Regularized partial correlations reveal such phenomena whereas regular sample correlations cannot.

The average partial correlations revealed by the sparse+latent estimator at all 31 sites were about 5 times lower than the sample correlations and less variable across sites (Fig. 6 A). 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. 6 B and C). However, there was an inverse relationship between the number of latent units and the average node degree (Fig. 6 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 differences in orientation preference and physical distances (lateral and by depth) between pairs of cells to the average sample correlations, average regularized partial correlations, and inferred connectivity between them (Fig. 7). The average partial correlations fell more rapidly with difference in preferred orientation (Fig. 7A and D) and lateral displacements at equal depths ($\pm 25 \mu\text{m}$) (Fig. 7B and E), and differences in depth at small ($\pm 25 \mu\text{m}$) lateral displacements (Fig. 7C and F).[31]

The connectivity expressed by the partial correlations in the regularized estimate had different organization for positive and negative interactions. Positive interactions fell rapidly as a function of difference in preferred orientation (Fig. 7D), lateral displacements (Fig. 7E), and displacement in depth (Fig. 7F). Negative interactions were much less selective (Fig. 7D–F).

Discussion

Functional and anatomical connectivity Systems neuroscience is the quest to discover principles whereby anatomical organization of neural circuits gives rise to their function. This line of inquiry requires expressive descriptions of the observed activity of neural circuits.[32]

The term *functional architecture* most commonly refers to stimulus response properties of neurons [52]. Over the past 50 years, tremendous progress has been made in relating the functional architecture to the anatomical, synaptic wiring of neuronal populations and individual neurons. This line of inquiry is based on the view of the brain as a complex filter of external stimuli.

In contrast to functional architecture, *functional connectivity* commonly refers to the degree of correlation, synchronization, or other forms of statistical association between neuronal populations or individual neurons inferred from spontaneous activity or trial-to-trial variations, not directly driven by external stimuli. Functional connectivity is thought to reflect internal, recurrent connectivity in neural circuits. However, less experimental evidence has been garnered to for links between functional and anatomical connectivity.

Neural correlations boast a rich history as indicators of functional connectivity. When measured on temporal scales of milliseconds, sharp peaks or troughs in the cross-correlogram between pairs of neurons

Figure 5 (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 estimate show more pronounced differences (**B** and **E**). **C.** Furthermore, the partial correlation matrix of the regularized estimate is decomposed into a sparse component with 82.2% off-diagonal zeros (bottom-left) and low-rank component of rank 15 (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 82.2% of cells pairs, equal to the fraction of zeros in the sparse partial correlation matrix. This interval contained 58.9% of the partial correlations. **G.** The 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.

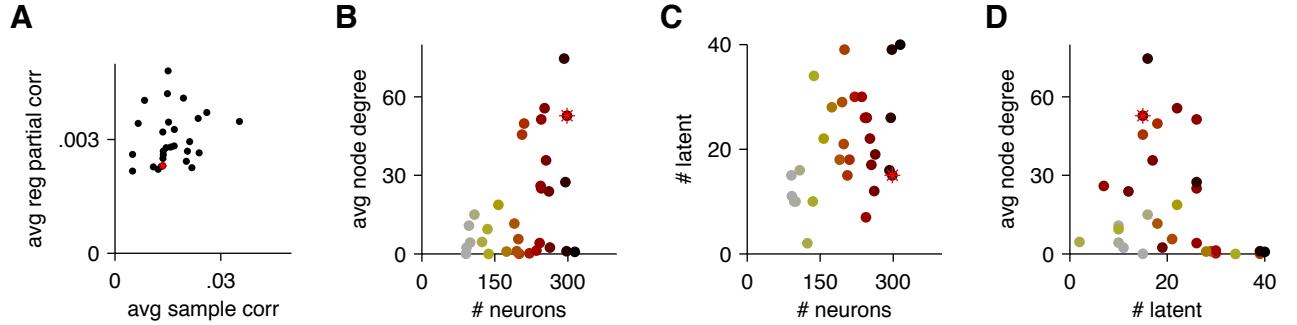


Figure 6. Properties of sparse+low-rank regularized estimates from all imaged sites

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 3 and 5. **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.

at latencies typical of synaptic transmission provide evidence of monosynaptic connection between pairs of cells [12, 27, 53] (although other processes can also generate sharp peaks in the cross-correlogram while many monosynaptic connections may not produce such signatures). Without targeted stimulation of specific cells or measurements of subthreshold intracellular events, such analysis remains the best method for detecting probable monosynaptic connections *in vivo*.

When measured at longer temporal scales, neural correlations reflect activations involving large ensembles of neurons and may reflect direct causal anatomical interactions, indirect causal interactions, or complex network-wide emergent features of activity with little correspondence to the local anatomical connectivity. Calcium imaging allows simultaneous measurements of the activity of large populations of cells *in vivo* [48, 49] and even of entire nervous systems [54, 55], but its temporal bandwidth is limited to the scale of tens or hundreds milliseconds by the kinetics of calcium-sensitive dyes, limiting its use to slower temporal scales at which correspondence between functional and anatomical connectivity is less direct.

Extraction of the correlation structure The word ‘structure’ implies a succinct, simplified representation that captures only the essential, significant features. Several forms of correlations structure can be extracted from the same data and used to describe functional connectivity in different ways.

Perhaps the least sophisticated yet common approach for extracting the structure of correlations is to threshold above a selected level of significance. Part of the appeal of thresholded correlations is their representation in the form of graphs that can be related to graphs of synaptic connectivity and made amenable to network analysis [56]. For example, [51] analyzed the graph theoretical properties of networks of thresholded correlations in development.

Alternatively, the correlation structure may be extracted in the form of low-rank approximations computed by principal component analysis, factor analysis, or activation maps conditioned on the activity of a single neuron. Low-rank components show synchronous multineuronal activation patterns without revealing graphs of interactions between neurons. They are often interpreted as emergent patterns of large circuits or brain state fluctuations [16].

Another reduced representation of correlation matrices is obtained by sparsifying its inverse, resulting in a graph of partial correlations. Under the assumption of a multivariate normal distribution or, more generally, an elliptical distribution, sparse partial correlations suggest a sparse network of dependencies,

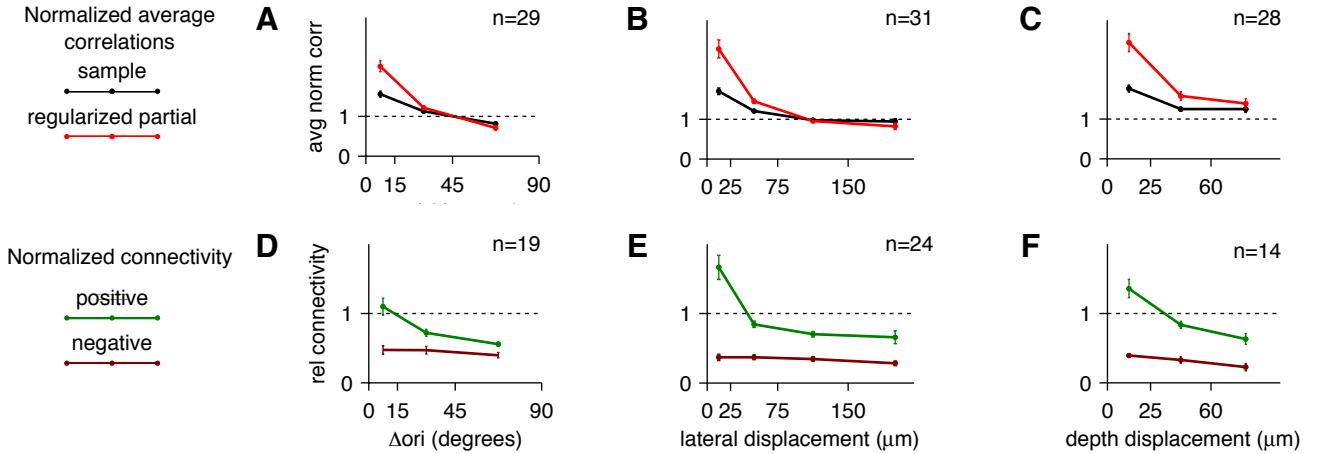


Figure 7. Dependence of correlations and partial correlations on orientation tuning differences and physical distance between cell pairs. **A–C.** Average partial correlations (red) estimated by $C_{\text{sparse+latent}}$ and average sample correlations (black) averaged across multiple imaged sites. In each site, the correlations were normalized by the respective average correlation shown in Fig. 6 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 μm, 25–75 μm, 75–150 μm, and 150+ μm. **C.** Average correlations between pairs of neurons displaced laterally by less than 25 μm separated in depth by distances in the intervals of 0–25 μm, 25–60 μm, and 60+ μm. **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.

i.e. a probabilistic graphical model. This property has led some investigators to propose that sparse inverse approximation of correlations may serve as closer estimates of direct anatomical interactions between brain regions in imaging studies than correlations [34, 35]. However, since deviations from elliptical distribution disrupt the correspondence between partial correlations and linear dependencies [37], it is not clear whether this is indeed the case. Partial correlations are meaningful when most interacting variables are observed. Perhaps for this reason, until now, partial correlations had not been considered in analysis of multineuronal recordings.

In this study, we sought to uncover the optimal low-dimensional approximation of the covariance matrix for efficient estimation of the true covariance matrices in cortical microcircuits on the temporal scale of 100–200 ms corresponding to the temporal resolution of spike rates from somatic calcium signals [49]. We found that networks of partial correlations were more statistically efficient than low-rank approximations of correlation matrices (Fig. S1 E). Furthermore, estimation became still more efficient when low-rank components were allowed to be combined with sparse partial correlations. We propose that correlation structures that are supported by the data are more likely to reflect anatomical, causal interactions. The network of sparse partial correlations suggests interactions between specific pairs whereas

the low-rank represents emergent activity patterns involving collective activations of larger circuits or global fluctuations. The revealed network of partial interactions was substantially different from that revealed by thresholding correlations (Fig. 5 C, F, H, I). In particular, it shows that many low correlations correspond to strong negative interactions (Fig. 5 F). Intuitively, this suggests that many low interactions are as surprising or significant considering the common correlations with other neurons. The analysis of sample correlations does not ascribe significance to correlation coefficients that are surprisingly low, based on their common correlations to other cells.

Functional connectivity from probabilistic models [33] Increasingly, functional connectivity is inferred by fitting probabilistic models such as various forms of multivariate Gaussian models, linear-nonlinear models, Generalized Linear Models [57], and maximum entropy models [38, 39, 58, 59]. In probabilistic models, functional dependencies are defined by multiplicative terms comprising the model. If the model can be decomposed as a product of terms so that a pair of variables can be segregated into separate terms, the two variables are conditionally independent: the activity of one neuron can be predicted from other neurons just as well with or without considering the second neuron, suggesting a lack of a direct causal interaction between them. Probabilistic models allow inferring latent variables to account for the activity of unobserved circuitry. [34]

Model selection Various estimation methods infer very different structures of functional connectivity from the same data. What kind of structure expresses most relevant information? How does one determine which functional connectivity structure is best supported by the empirical data? The field of machine learning offers a rich toolbox of model selection criteria. Parametric criteria such as the *Bayesian model selection criterion*, *Akaike Information criterion*[35], and *Bayes (or Schwartz) Information Criterion* (BIC) are computationally efficient but rely on generally unsupportable assumptions[36] about the data generating process. The most assumption-free yet more computationally intensive are procedures based on cross-validation [60] and, when models have hyperparameters, nested cross-validation to select their values. It is commonly assumed that probabilistic models that demonstrate the highest average cross-validated log likelihood yield a better representation of the functional connectivity. For example, [61] demonstrated that their proposed Reliable Network Model (a maximum-entropy model on the multivariate binary domain constrained by a sparse heuristically selected network of correlations of several orders) yielded higher cross-validated log-likelihoods than the full-pairwise Ising model. The authors use this fact to support the claim that the former model revealed a more anatomically relevant structure of functional connectivity than the pairwise Ising model.

However, in simulation, we showed that, more constrained, simpler models often outperformed complex models in cross-validation studies even when their structure was differed from the true structure (Fig. 2 row 5), only because these more constrained models were less sensitive to sampling noise. [37] Therefore, by itself, better cross-validation performance of a model does imply closer correspondence of its structure to the functional structure of the data generating process. For example, a properly sparsified pairwise Ising model could outperform the Reliable Network model. Unfortunately, optimal sparsification of Ising models is no computationally tractable procedure exists to infer the optimal sparse structure of dependencies in an Ising model.[38]

In our study we selected between families of models based on their cross-validation performance. The winning model family was more flexible than the others and therefore more susceptible to noise.[39] The fact that it still performed better indicated that the additional features were required. We conclude that under these circumstances, the selected model can be asserted to better represent the true structure of the data generating process, at least in comparison to the other three models.

In our study, we focused on estimating a specific parameter of the distribution (its covariance matrix) rather than fitting a specific parametric distribution. Under the loss function (Eq. 9) we selected, the covariance matrix estimation methods are equivalent to fitting a multivariate Gaussian model and select-

ing the best structure on the basis of cross-validated log-likelihood. However, the ranking of models was preserved under other loss functions (Fig. S??), indicating that model selection results were not strongly dependent on the assumption of gaussianity. We defend the practice of evaluating and optimizing probabilistic models with respect to their parameters that are more relevant for functional connectivity rather than maximizing their cross-validated log likelihoods, which are sensitive to all parameters. [40]For example, model A may attain higher cross-validated log likelihood than model B due to its better ability to capture the tails of the distribution while model B better estimates the correlation matrix. We maintained that model B must be more relevant for studies of functional connectivity because correlations do reflect aspects of functional connectivity. [41]

The correlation structures we considered also corresponded to structures of conditional dependencies in multivariate Gaussian models or, more generally, elliptical distributions. For example, zero partial correlation indicates conditional independence between a pair of variables under an elliptical distribution, but this correspondence weakens or breaks down with deviation from elliptical distributions [37]. Other probabilistic models, when optimally sparsified for best cross-validated estimation of the covariance matrix, could suggest very different forms of low-dimensional representations (structures) of the covariance matrix. For example, spatiotemporal Ising models, marginalized to reproduce the correlation matrix at the same temporal scale, would suggest an alternative correlation structure from the same data. [42]At this time we do not have computationally efficient methods for finding sparse structures of dependencies in Ising models optimized with respect to cross-validated estimation of their cross-correlation matrices. If we did, and the estimator thus constructed outperformed our $C_{\text{sparse+lowrank}}$ estimator, we would be adopt this new estimator as one producing the best structure of the connectivity. Various forms of GLMs could also be evaluated with in the same way: sparsified for best estimation of the covariance matrix (not log-likelihood) and compared to the estimators studied in this study. Again, if this new covariance estimator outperforms, $C_{\text{sparse+lowrank}}$, we will be compelled to adapt it for estimating the functional connectivity.

Big neural data Transformational breakthroughs in systems neuroscience are anticipated to follow the ongoing technological advances that will allow simultaneous recordings of the spiking activity of much larger subsets of neurons in functioning neural circuits than have been possible thus far [62]. It is far from certain whether massively multineuronal recordings will readily reveal new principles of neural function by application of current statistical techniques. Large datasets aggravate the numerical problems and the model selection caveats described in this paper. The number of parameters in models of population activity grows quadratically. Models of joint population activity feature large numbers of parameters, that growing quadratically or faster with population size. Without ways to restrict such models, they quickly become ill conditioned (e.g. [63]). Statistical models of population activity will need to have both correct mathematical forms of interactions and optimal ways of restricting them so that the total number of estimated parameters remains commensurate with the amount of measurements.

Our study employs regularization, which restricts the models based on the data itself. Other models may use restrictions that are based on various heuristics in addition to regularization. These methods have numerous caveats when it comes to inferring the optimal structure of functional connectivity from such methods. Improved estimation of covariances and the general framework of producing such estimators for specific neural populations can be applied to various circuits and serve as a statistical description of their functional organization.

Correspondence to anatomical organization If the ultimate goal of this line of inquiry to find descriptions of functional connectivity that are most informative about the anatomical and physiological organization of the circuit, then the models must be directly validated according to their correspondence to the anatomical organization. Little empirical evidence presently exists to show which models of connectivity, when fitted to the recordings of activity, reveal functional connectivity with best correspondence to anatomical connectivity.

In this study we find that the structure of partial correlation of the optimal estimator had stronger dependencies on the separation in depth, lateral position, and orientation preferences than sample correlations (Fig. 7 A–C). The inferred positive and negative partial correlations too had different structures COMPARED TO NOISE CORRELATIONS ?? [43](Fig. 7 D–F). Also the average partial correlations were more consistent between imaged sites than average sample correlations (Fig. 6 A). Although these effects could conceivably results from trivial numerical artifacts, they are consistent with improved representation of functional connectivity by the proposed estimator.

Additional studies with additional markers of circuit architecture such as cell types and synaptic connectivity could shed additional light on the ability of models to reveal meaningful functional connectivity. For example, labeled interneurons could be studied for their distinct connectivity patterns contrasted with pyramidal cells. I THIKN WE NEED ONE LAST SENTNECE HERE TO SAY THAT FUNCTIONAL CONNECTIVITY IF IT RELATED TO ANALTOMICAL IS IMPORTANT SINCE IT CAN TELL US HOW CIRCUITS INTERACT TO COMPUTE MENTION EXAMPLE OF STATES, SLEEP HERE ETC. I.E. GROUPS CELLS MAY BE CONNECTED BUT ONLY INTERACTING IN STATE A AND NOT B.

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 [49]. Briefly, C57BL/6J mice (aged p40–60) were used. Anesthesia was initiated with isoflurane (3

Calcium imaging began 1 hour after dye injection. All imaging was performed using the 3D-RAMP two-photon microscope described in [49]. 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/m², 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 map using a pseudo-random sequence of drifting gratings at sixteen equally spaced directional 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 dataset) 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 (<https://github.com/datajoint/>) 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 artifacts 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 [50].

Orientation tuning was computed by fitting the mean firing rates in response to gratings of directions ϕ 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 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 -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 mins. Only traces whose average deconvolved signal was greater than 1% of the median of the recorded population 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

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. 2 rows 3 and 4) and to analyze the partial correlation structure of the sparse+latent estimator (Fig. 5–7).

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_{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:

$$D = (1 - \alpha)C_0 \circ I + \alpha \frac{1}{p} \text{tr}(C_0)I \quad (14)$$

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 [64], 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_{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 Ψ were computed by solving the minimization problem

$$(L, \Psi) = \arg \min_{\tilde{L}, \tilde{\Psi}} \mathcal{L}(\tilde{L}\tilde{L}^\top + \tilde{\Psi}, C_0), \quad (15)$$

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 Ψ under the multivariate Gaussian distribution. The final regularized estimate is obtained by linear shrinkage toward the factor model (Eq. 6).

Estimator C_{sparse} has one hyperparameter λ , 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.

$$Z = \arg \min_{\tilde{Z} \succ 0} \mathcal{L}(\tilde{Z}^{-1}, C_0) + \lambda \|\tilde{Z}\|_1 \quad (16)$$

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* [65]. To solve this minimization problem, we modified the alternative-direction method of multipliers (ADMM) algorithm developed by [44]. Then, after the zero structure was determined, the remaining coefficients were fitted without penalty:

$$S = \arg \min_{\tilde{S} \in Z^\#} \mathcal{L}(\tilde{S}, C_0), \quad (17)$$

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

$$S^* = \begin{pmatrix} S & S_{12} \\ S_{12}^\top & S_{22} \end{pmatrix}, \quad (18)$$

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

$$C_{\text{sparse+lowrank}} = (S^*)^{-1} = (S - S_{12}S_{22}^{-1}S_{12}^\top)^{-1} \quad (19)$$

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 a single matrix LL^\top of rank d , where L is of size $p \times d$. 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 [43, 44]:

$$(Z, \cdot) = \arg \min_{\tilde{Z}, \tilde{L}} \mathcal{L}(\tilde{Z} - \tilde{L}\tilde{L}^\top) + \|\tilde{Z}\|_1 \quad (20)$$

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

$$(S, L) = \arg \min_{\tilde{S} \in Z^\#, \tilde{L}} \mathcal{L}(\tilde{S} - \tilde{L}\tilde{L}^\top) \quad (21)$$

The partial correlation matrix computed from the overall estimate, computed from $C_{\text{sparse+lowrank}}$ 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. 5 B and E and Fig. 6 A, and Fig. 7 A–C). For analyses that required a sparse network (Fig. 5 F, G, H and Fig. 6 B, D and Fig. 7 D–F) was computed using the extended precision matrix S^* :

$$P_{\text{sparse}} = (S \circ I)^{-\frac{1}{2}} S (S \circ I)^{-\frac{1}{2}} \quad (22)$$

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. 2 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. Josic K, Shea-Brown E, Doiron B, de la Rocha J (2009) Stimulus-dependent correlations and population codes. *Neural computation* 21: 2774–2804.
9. 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.
10. 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.
11. Smith MA, Kohn A (2008) Spatial and temporal scales of neuronal correlation in primary visual cortex. *J Neurosci* 28: 12591–12603.
12. 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* .
13. 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.
14. 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.
15. Chiu C, Weliky M (2002) Relationship of correlated spontaneous activity to functional ocular dominance columns in the developing visual cortex. *Neuron* 35: 1123–1134.
16. Kenet T, Bibitchkov D, Tsodyks M, Grinvald A, Arieli A (2003) Spontaneously emerging cortical representations of visual attributes. *Nature* 425: 954–956.

17. Cohen MR, Newsome WT (2008) Context-dependent changes in functional circuitry in visual area mt. *Neuron* 60: 162–173.
18. Cohen MR, Maunsell JH (2009) Attention improves performance primarily by reducing interneuronal correlations. *Nature neuroscience* 12: 1594–1600.
19. Hansen BJ, Chelaru MI, Dragoi V (2012) Correlated variability in laminar cortical circuits. *Neuron* 76: 590–602.
20. Smith MA, Jia X, Zandvakili A, Kohn A (2013) Laminar dependence of neuronal correlations in visual cortex. *Journal of neurophysiology* 109: 940–947.
21. 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.
22. 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.
23. Goard M, Dan Y (2009) Basal forebrain activation enhances cortical coding of natural scenes. *Nat Neurosci* 12: 1444–9.
24. Kohn A, Zandvakili A, Smith MA (2009) Correlations and brain states: from electrophysiology to functional imaging. *Curr Opin Neurobiol* 19: 434–8.
25. 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.
26. Perkel DH, Gerstein GL, Moore GP (1967) Neuronal spike trains and stochastic point processes: II. simultaneous spike trains. *Biophysical journal* 7: 419–440.
27. Moore GP, Segundo JP, Perkel DH, Levitan H (1970) Statistical signs of synaptic interaction in neurons. *Biophysical Journal* 10: 876–900.
28. Shadlen MN, Newsome WT (1998) The variable discharge of cortical neurons: implications for connectivity, computation, and information coding. *J Neurosci* 18: 3870–96.
29. Salinas E, Sejnowski TJ (2001) Correlated neuronal activity and the flow of neural information. *Nature Reviews Neuroscience* 2: 539–550.
30. 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.
31. Rosenbaum R, Josić K (2011) Mechanisms that modulate the transfer of spiking correlations. *Neural computation* 23: 1261–1305.
32. 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.
33. Peng J, Wang P, Zhou N, Zhu J (2009) Partial correlation estimation by joint sparse regression models. *Journal of the American Statistical Association* 104.
34. 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.

35. 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.
36. Koller D, Friedman N (2009) Probabilistic graphical models: principles and techniques. MIT press.
37. Loh PL, Wainwright MJ (2012) Structure estimation for discrete graphical models: Generalized covariance matrices and their inverses. arXiv preprint arXiv:12120478 .
38. 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.
39. Tkacik G, Schneidman E, Berry I, Michael J, Bialek W (2006) Ising models for networks of real neurons. arXiv preprint q-bio/0611072 .
40. Ledoit O, Wolf M (2004) A well-conditioned estimator for large-dimensional covariance matrices. *Journal of multivariate analysis* 88: 365–411.
41. Bickel PJ, Li B, Tsybakov AB, van de Geer SA, Yu B, et al. (2006) Regularization in statistics. *Test* 15: 271–344.
42. Dempster A (1972) Covariance selection. *Biometrics* : 157–175.
43. 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.
44. Ma S, Xue L, Zou H (2013) Alternating direction methods for latent variable gaussian graphical model selection. *Neural computation* : 1–27.
45. 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.
46. Fan J, Fan Y, Lv J (2008) High dimensional covariance matrix estimation using a factor model. *Journal of Econometrics* 147: 186–197.
47. Reddy GD, Saggau P (2005) Fast three-dimensional laser scanning scheme using acousto-optic deflectors. *J Biomed Opt* 10: 064038.
48. 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* .
49. Cotton RJ, Froudarakis E, Storer P, Saggau P, Tolias AS (2013) Three-dimensional mapping of microcircuit correlation structure. *Frontiers in Neural Circuits* 7: 151.
50. 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.
51. Malmersjö S, Rebello 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.
52. Reid RC (2012) From functional architecture to functional connectomics. *Neuron* 75: 209–217.

53. Alonso JM, Martinez LM (1998) Functional connectivity between simple cells and complex cells in cat striate cortex. *Nature neuroscience* 1: 395–403.
54. Leung LC, Wang GX, Mourrain P (2013) Imaging zebrafish neural circuitry from whole brain to synapse. *Frontiers in neural circuits* 7.
55. 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.
56. Bullmore E, Sporns O (2012) The economy of brain network organization. *Nature Reviews Neuroscience* 13: 336–349.
57. 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.
58. 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.
59. 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.
60. Arlot S, Celisse A (2010) A survey of cross-validation procedures for model selection. *Statistics Surveys* 4: 40–79.
61. 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.
62. Alivisatos AP, Chun M, Church GM, Deisseroth K, Donoghue JP, et al. (2013) The brain activity map. *Science* 339: 1284–1285.
63. 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.
64. Schäfer J, Opgen-Rhein R, Strimmer K (2010) corpcor: Efficient estimation of covariance and (partial) correlation. *R package version* 15 7.
65. Friedman J, Hastie T, Tibshirani R (2008) Sparse inverse covariance estimation with the graphical lasso. *Biostatistics* 9: 432–441.

Supporting Information

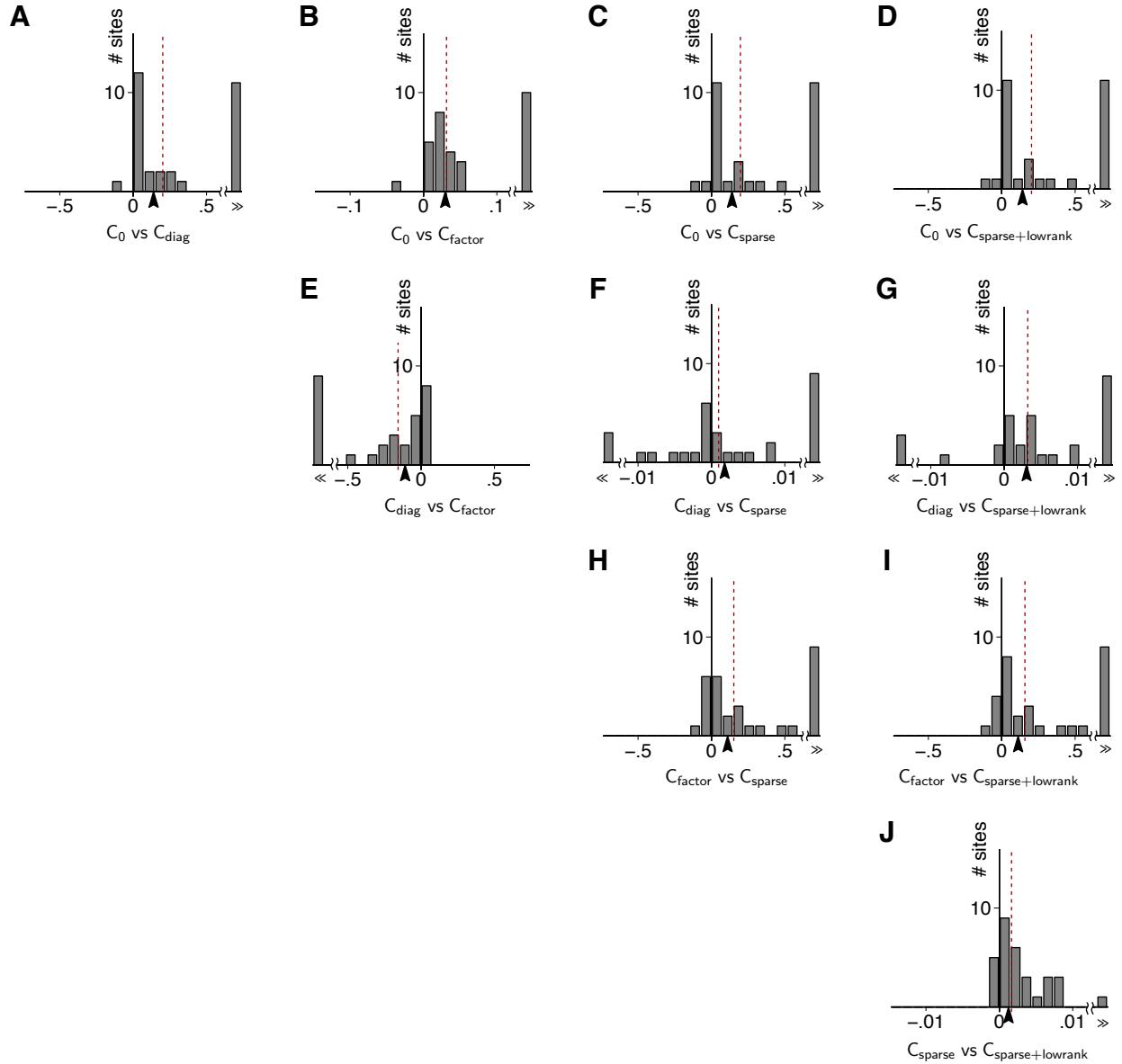


Figure 1. All-to-all performance comparisons of the sample covariance matrix and the four regularized estimators with respect to multivariate normal cross-validation loss.

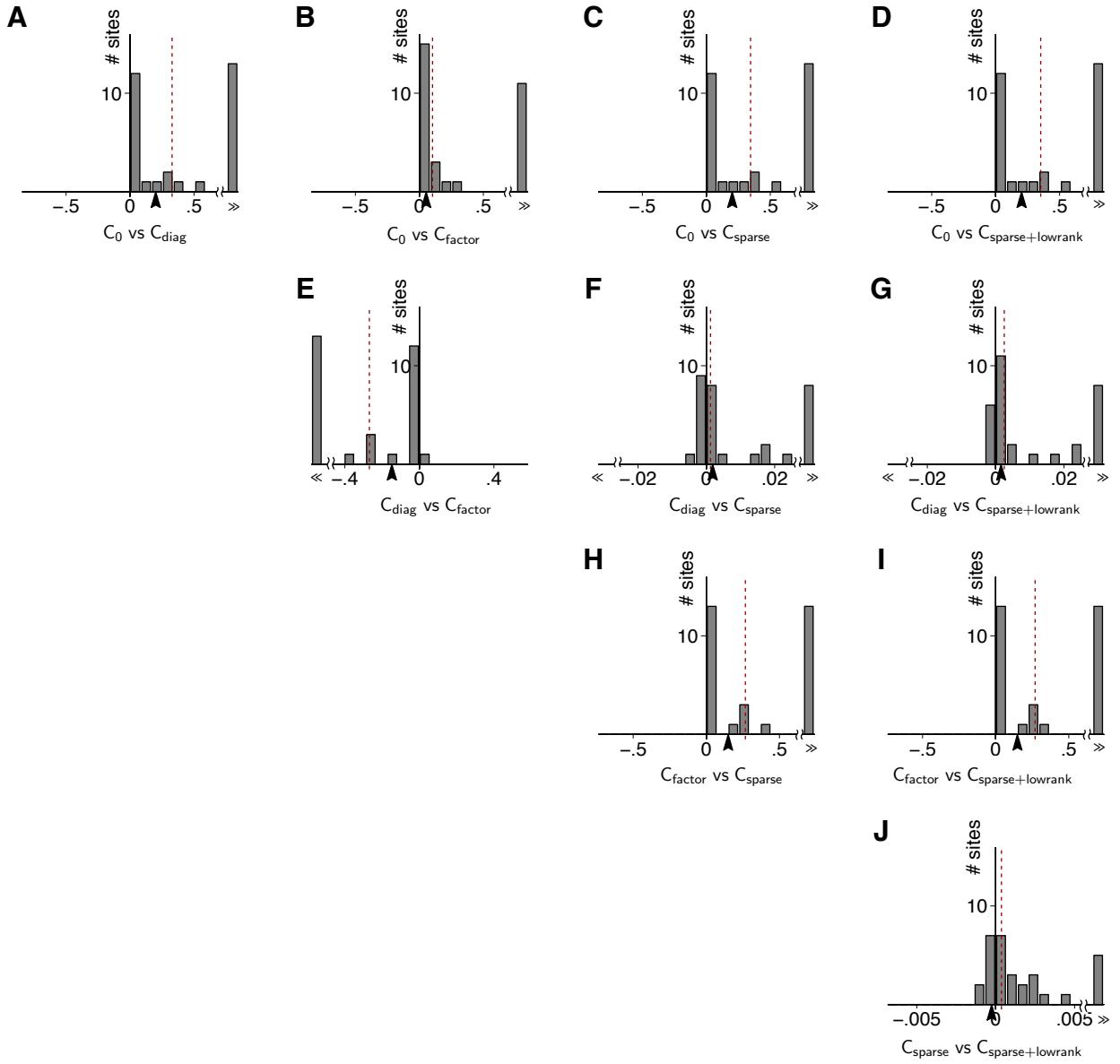


Figure 2. All-to-all performance comparisons of the sample covariance matrix and the four regularized estimators with respect to quadratic cross-validation loss.