

Improved estimation and interpretation of correlations in neural circuits

Dimitri Yatsenko¹, 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: atolias@cns.bcm.edu**

Abstract

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

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

Pearson correlations between the spiking activity of pairs of neurons are among the most familiar measures of functional connectivity [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 aspects of synaptic connectivity [12]. Interest in neural correlations has been sustained by a series of discoveries of their nontrivial relationships to various aspects of circuit organization such as the physical distances between the neurons [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 [26–28], changes due to sensory stimulation and global brain states [5, 21, 29–32].

Neural correlations do not come with ready or unambiguous mechanistic interpretations. They can arise from monosynaptic or polysynaptic interactions, common or correlated inputs, oscillations, top-down modulation, and background network fluctuations, and other mechanisms [33–38]. But multineuronal recordings do provide more information than an equivalent number of separately recorded pairs of cells. For example, the eigenvalue decomposition of the covariance matrix expresses shared correlated activity components across the population; common fluctuations of population activity may be accurately represented by only a few eigenvectors that affect all correlation coefficients. On the other hand, a correlation matrix can be specified using the *partial correlations* between pairs of the recorded neurons. The partial correlation coefficient between two neurons reflects their linear association conditioned on the activity of all the other recorded cells [39]. Under some assumptions, partial correlations measure conditional independence between variables and may more directly approximate causal effects between components of complex systems than correlations [39]. For this reason, partial correlations have been used to describe interactions between genes in functional genomics [40, 41] and between brain regions in imaging studies [42, 43]. These opportunities have not yet been explored in neurophysiological studies where most analyses have only considered the distributions of pairwise correlations [2, 4, 5, 13].

However, estimation of correlation 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 the largest eigenvalues) [44], and poorly conditioned partial correlations [40]. The *sample correlation matrix* is an unbiased estimate of the true correlations but its many free parameters make it sensitive to sampling noise. As a result, on average, the sample correlation matrix is farther from the true correlation matrix than some structured estimates.

Estimation can be improved through *regularization*, the technique of deliberately imposing a structure on an estimate in order to reduce its estimation error [40, 45]. To ‘impose a structure’ on an estimate

means to bias (‘shrink’) it toward a reduced representation with fewer free parameters, the *target estimate*. The optimal target estimate and the optimal amount of shrinkage can be obtained from the data sample either analytically [40, 44, 46] or by cross-validation [47]. An estimator that produces estimates that are, on average, closer to the truth for a given sample size is said to be more *efficient* than other estimators.

Although regularized covariance matrix estimation is commonplace in finance [46], functional genomics [40], and brain imaging [43], surprisingly little work has been done to identify optimal regularization of neural correlation matrices.

Improved estimation of the correlation matrix is beneficial in itself. For example, improved estimates can be used to optimize decoding of the population activity [47]. But reduced estimation error is not the only benefit of regularization. Finding the most efficient among many regularized estimators leads to insights about the system itself: the structure of the most efficient estimator is a parsimonious representation of the regularities in the data.

The advantages due to regularization increase with the size of the recorded population. With the advent of big neural data [48], the search for optimal regularization schemes will become increasingly relevant in any model of population activity. Since optimal regularization schemes are specific to systems under investigation, the inference of functional connectivity in large-scale neural data will entail the search for optimal regularization schemes. Such schemes may involve combinations of heuristic rules and numerical techniques specially designed for given types of neural circuits.

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

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

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

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

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

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

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

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

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

Results

Covariance estimation

The covariance matrix is defined as

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

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

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

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

where n is the number of time bins in T , yields an unbiased estimate of the covariance matrix so that $\mathbb{E} [C_{\text{sample}}] = \Sigma$.

When the mean is estimated from the same sample, C_{sample} becomes biased toward zero. However, in all the cases when the unbiasedness is required in our study, the mean will be estimated from a separate sample.

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

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

where \circ denotes the entrywise matrix product (Hadamard product) and I is the $p \times p$ identity matrix.

The *partial correlation* between a pair of variables is the Pearson correlation coefficient of the residuals of the linear least-squares predictor of their activity based on all the other variables, excluding the pair [39,49]. Partial correlations figure prominently in probabilistic *graphical modeling* wherein the joint distribution is explained by sets of two-way interactions [39]. For the multivariate Gaussian distribution, zero partial correlations indicate conditional independence of the pair, implying a lack of direct interaction [39,50]. More generally, partial correlations can serve as a measure of conditional independence under the assumption that most dependencies in the system include strong linear effects [39,51]. As neural recordings become increasingly dense, partial correlations may prove useful as indicators of conditional independence (lack of functional connectivity) between pairs of neurons.

Pairwise partial correlations are closely related to the elements of the *precision matrix*, *i.e.* the inverse of the covariance matrix [39,50]. Zero elements in the precision matrix signify zero partial correlation between the two variables. Given the covariance estimate C , the matrix of partial correlations P is computed by normalizing the rows and columns of the *precision matrix* C^{-1} to produce negative unit entries on the diagonal:

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

As the size of recorded neuronal populations increases, without substantial increases in recording durations, the *condition number* of the sample covariance matrix also increases, making the partial correlation estimates *ill-conditioned*: small errors in covariance estimates translate into much greater errors in partial correlations after the inversion. With massively multineuronal recordings, partial correlations cannot be estimated without *regularization* [40,44].

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

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

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

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

Estimator C_{factor} approximates the covariance matrix by the factor model $L + D$, where L is the $p \times p$ positive semidefinite matrix of a low rank and D is a diagonal matrix of independent variances. This approximation is the basis for *factor analysis* [49]. The estimator,

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

has two hyperparameters: the rank of L (*i.e.* the number of latent factors) and the shrinkage intensity λ to shrink the independent variances toward their mean value \bar{D} . The structure imposed by C_{factor} favors conditions in which the population activity is linearly driven by a number of latent factors that affect many cells while direct interactions between the recorded cells are insignificant (Fig. 1 Row 1, B).

Estimator C_{sparse} is produced by approximating the sample covariance matrix by the inverse of a sparse matrix S :

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

Here S is a sparse matrix, *i.e.* one in which a large fraction of off-diagonal elements are set to zero. The estimator has one hyperparameter that determines the sparsity (fraction of off-diagonal zeros) in S . The problem of finding the optimal set of non-zero elements of the precision matrix is known as *covariance selection* [50]. The structure imposed by C_{sparse} favors conditions in which neural correlations arise from direct linear effects ('interactions') between some pairs of neurons (Fig. 1 Row 1, C).

Estimator $C_{\text{sparse+latent}}$ is obtained 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 . See Methods for a more detailed explanation. The structure imposed by $C_{\text{sparse+latent}}$ favors conditions in which the activity of neurons is determined by linear effects between some observed pairs of neurons and linear effects from several latent units (Fig. 1 Row 1, D) [52, 53].

We refer to the sparse partial correlations in estimators C_{sparse} and $C_{\text{sparse+latent}}$ as 'interactions'.

Simulation

We next demonstrate how the most efficient among different regularized estimators can reveal the structure of correlations. We constructed four families of 50×50 covariance matrices, each with structure that matched one of the four regularized estimators (Fig. 1 Row 2, A–D and Methods). We used these covariance matrices as the ground truth in multivariate normal distributions with zero means and drew samples of various sizes. Even with relatively large sample sizes (*e.g.* $n = 500$), the sample correlation matrices were contaminated by sampling noise and their underlying structures were difficult to discern (Fig. 1 Row 3).

The evaluation of any covariance matrix estimator, C , is performed with respect to a *loss function* $\mathcal{L}(C, \Sigma)$ to quantify its discrepancy from truth, Σ . 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 is related to the multivariate normal log-likelihood function $L(\Sigma|C_{\text{sample}})$,

$$\mathcal{L}(X, Y) \equiv -\ln(2\pi) - \frac{2}{pn} L(X|Y) \quad (10)$$

Normalization by p and n makes the values of the loss function comparable across different population sizes and sample sizes.

Although the two functions are similar in form, they differ conceptually: The log-likelihood $L(\Sigma|C_{\text{sample}})$ is a function of the parameter Σ given the sample covariance matrix C_{sample} whereas the loss function $\mathcal{L}(C, \Sigma)$ expresses the deviation of an arbitrary estimate C from the truth Σ . Minimum expected loss may be found at a point that is far removed from the point of maximum likelihood given the same data.

The choice of the normal loss function is motivated, in part, 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 [40, 44, 54, 55].

We drew 30 independent samples of sizes $n = 250, 500, 1000, 2000$, and 4000 from each model and computed the *excess loss* $\mathcal{L}(C, \Sigma) - \mathcal{L}(\Sigma, \Sigma)$ for each of the five estimators. The hyperparameters of the regularized estimators were optimized by nested cross-validation using only the data in the sample. All the regularized estimators produced better estimates (lower excess loss) than the sample covariance matrix. However, estimators whose structure matched the true model outperformed the other estimators (Fig. 1 Row 5). Note that when the ground truth had zero correlations (Column A), C_{factor} performed equally well to C_{diag} because it correctly identified zero factors and only estimated the individual variances. Similarly, when the number of latent units was zero (Column C), $C_{\text{sparse+latent}}$ performed equally well to C_{sparse} because it correctly identified zero latent units.

With increasing sample sizes, all estimators converged to the truth (zero excess loss). However, the discriminability between their performances increased with sample size (data not shown).

With empirical data, the ground truth, Σ , is not accessible and excess loss cannot be computed directly. Instead, the loss function may be estimated from the sample through *validation*. In a validation procedure, an independent *testing sample* is set aside to compute an additional sample covariance estimate C'_{sample} to validate the estimate C computed from the *training sample*. The estimate C'_{sample} is used to compute the *validation loss* $\mathcal{L}(C, C'_{\text{sample}})$, which approximates the loss $\mathcal{L}(C, \Sigma)$.

The normal loss (Eq. 9) is particularly suitable because it is additive in its second argument, *i.e.*

$$\mathcal{L}(C, Z_1) + \mathcal{L}(C, Z_2) \equiv \mathcal{L}(C, Z_1 + Z_2)$$

Then, with fixed C , validation loss is an unbiased estimate of true loss:

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

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 and validation does not exactly evaluate a discrepancy from truth.

The loss function above assumes a single covariance matrix across all conditions in the dataset. However, neuroscientists often estimate a common correlation matrix across multiple stimulus conditions when the variances of responses are conditioned on the stimulus [56, 57]. In this case, the stimulus-dependent means and variances are estimated from the training dataset along with a common correlation matrix. An adaptation of the normal loss function to this case is described in Methods.

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 testing sample with the other $K - 1$ subsets serving as the training sample. The validation losses from all such ‘folds’ are averaged to produce the *cross-validation loss*.

Cross-validation loss accurately reproduced the relationships between the excess losses of the estimators (Fig. 1 Row 6), confirming that cross-validation can evaluate true loss.

This simulation study demonstrated that, with sufficiently large sample sizes, imposing the correct type of structure on the estimate leads to greater estimation improvement than imposing other structures.

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

We recorded the calcium activity of densely sampled populations of neurons in the supragranular layers in primary visual cortex of fentanyl-anesthetized mice using fast random-access 3D scanning two-photon microscopy during visual stimulation (Fig. 2 A–B) [58–60]. 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 and D). The firing rates were inferred using sparse nonnegative deconvolution [61] (Fig. 2 C). Only cells that produced detectable calcium activity were included in the analysis (see Methods). First, 30 repetitions of full-field drifting gratings of 16 directions were presented in random order. Each grating was displayed for 500 ms, without intervening blanks. This stimulus was used to compute the orientation tuning of the recorded cells (Fig. 2 D). To estimate the noise correlation matrix, we presented only 2 distinct directions in some experiments or 5 directions in others with 100–300 repetitions of each direction. Each grating lasted 1 second and was followed by a 1-second blank. The traces were then binned into 150 ms intervals aligned on the stimulus onset for the estimation of the correlation matrix. The sample correlation coefficients were largely positive and low (Fig. 2 E and F). The average value of the correlation coefficient across sites ranged from 0.0065 to 0.051 with the mean across sites of 0.018 (Fig. 5 D).

In these densely sampled populations, direct interactions between cells are likely to influence the patterns of population activity. We therefore hypothesized that covariance matrix estimators that explicitly modeled the partial correlations between pairs of neurons (C_{sparse} and $C_{\text{sparse+latent}}$) would have a performance advantage. However, the observed neurons must also be strongly influenced by global activity fluctuations and by unobserved common inputs to the advantage of estimators that explicitly model common fluctuations of the entire population: C_{factor} and $C_{\text{sparse+latent}}$. If both types of effects are significant, then $C_{\text{sparse+latent}}$ should outperform the other estimators.

To test this hypothesis, we computed the relative cross-validation loss of estimators C_{sample} , C_{diag} , C_{factor} , and C_{sparse} with respect to $C_{\text{sparse+latent}}$ in $n = 27$ imaged sites in 14 mice. The hyperparameters of each estimator were optimized by nested cross-validation (See Fig. S1 and Methods). Indeed, the sparse+latent estimator outperformed the other estimators (Fig. 3). The respective median differences of the cross-validation loss were 0.039, 0.0016, 0.0029, and 0.0059 nats/cell/bin, significantly greater than zero ($p < 0.01$ in each comparison, $n = 27$ sites in 14 mice, Wilcoxon signed rank test).

The same evaluation based on the quadratic loss function,

$$\mathcal{L}(C, C'_{\text{sample}}) = \frac{1}{p^2} \text{tr}(C^{-1}C'_{\text{sample}} - I)^2, \quad (11)$$

instead of the normal loss function reproduced the same relationship between the estimators (Fig. S2). This suggests that the results of this study are robust to the choice of the loss function and do not depend on the assumption of gaussianity.

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

We examined the composition of the $C_{\text{sparse+latent}}$ estimates at each imaged site (Fig. 4 and Fig. 5). Although the regularized estimates were similar to the sample correlation matrix (Fig. 4 A and B), the corresponding partial correlation matrices differed substantially (Fig. 4 C and D). The estimates separated two sources of correlations: a network of linear interactions expressed by the sparse component of the inverse and latent units expressed by the low-rank components of the inverse (Fig. 4 E). The sparse partial correlations revealed a network that differed substantially from the network composed of the greatest coefficients in the sample correlation matrix (Fig. 4 F, G, H, and I).

In the example site (Fig. 4), the sparse component had 92.8% sparsity (or conversely, 7.2% connectivity: $\text{connectivity} = 1 - \text{sparsity}$) with average node degree of 20.9 (Fig. 4 G). The average node degree, *i.e.* the average number of interactions linking each neuron, is related to connectivity as $\text{degree} = \text{connectivity} \cdot (p - 1)$, where p is the number of neurons. The low-rank component had rank 72, denoting 72 inferred latent units. The number of latent units increased with population size (Fig. 5 A) but the connectivity was highly variable (Fig. 5 B): Several sites, despite their large population sizes, were driven by latent units and had few pairwise interactions. This variability may be explained by differences in brain states and recording quality and warrants further investigation.

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

While the sample correlations were mostly positive, the sparse component of the partial correlations ('interactions') had a high fraction (28.7% in the example site) of negative values (Fig. 4 F). The fraction of negative interactions increased with the inferred connectivity (Fig. 5 D), suggesting that negative interactions can be inferred only after a sufficient density of positive interactions has been uncovered.

Thresholded sample correlations have been used in several studies to infer pairwise interactions [26, 62–64]. We therefore compared the interactions in the sparse component of $C_{\text{sparse+latent}}$ to those obtained from the sample correlations thresholded to the same level of connectivity. The networks revealed by the two methods differed substantially. In the example site with 7.2% connectivity in $C_{\text{sparse+latent}}$, only 27.7% of the connections coincided with the above-threshold sample correlations (Fig. 4 F, H, and I). In particular, most of the inferred negative interactions corresponded to low sample correlations (Fig. 4 F) where high correlations should be expected given the rest of the correlation matrix.

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

We examined how the structure of the $C_{\text{sparse+latent}}$ estimates related to the differences in orientation preference and to the physical distances separating pairs of cells (Fig. 6). Five sites with highest pairwise connectivities were included in the analysis. Partial correlations were computed using Eq. 4 based on the regularized estimate, including both the sparse and the latent component. Connectivity was computed as the fraction of pairs of cells connected by non-zero elements (interactions) in the sparse component of the estimate, distinguishing between the positive and negative connectivities.

First, we analyzed how correlations and connectivity depended on the difference in preferred orientations (Δori) of pairs of significantly ($\alpha = 0.05$) tuned cells. The partial correlations decayed more rapidly with Δori than did sample correlations ($p < 10^{-9}$ in each of the five sites, two-sample t -test of the difference of the linear regression coefficients). Positive connectivity decreased with Δori ($p < 0.005$ in each of the five sites, t -test on the logistic regression coefficient) whereas negative connectivity did not decrease (Fig. 6 D): The slope in the logistic model of connectivity with respect to Δori was significantly higher for positive than for negative interactions ($p < 0.04$ in each of the five sites, two-sample t -test of the difference of the

logistic regression coefficient).

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

Discussion

Functional connectivity as a network of pairwise interactions

Functional connectivity is often represented as a graph of pairwise interactions. The goal of many studies of functional connectivity has been to estimate anatomical connectivity from observed multineuronal spiking activity. For example, characteristic peaks and troughs in the pairwise cross-correlograms of recorded spike trains contain statistical signatures of directional monosynaptic connections and shared synaptic inputs [12, 14, 33, 34, 65]. Such signatures are ambiguous as they can arise from network effects other than direct synaptic connections [66]. With simultaneous recordings from more neurons, ambiguities can be resolved by inferring the conditional dependencies between pairs of neurons. Direct causal interactions between neurons produce statistical dependency between them even after conditioning on the state of the remainder of the network and external input. Therefore, conditional independence can signify the absence of a direct causal influence.

Conditional dependencies can be inferred by fitting a probabilistic model of the joint population activity. For example, generalized linear models (GLMs) have been constructed to include biophysically plausible synaptic integration, membrane kinetics, and individual neurons' stimulus drive [67]. Maximum entropy models constrained by observed pairwise correlations are among other models with pairwise coupling between cells [68–72]. Assuming that the population response follows a multivariate normal distribution, the conditional dependencies between pairs of neurons are expressed by the partial correlations between them. Each probabilistic model, fitted to the same data may reveal a completely different network of 'interactions', *i.e.* conditional dependencies between pairs of cells.

It is not yet clear which approach provides the best correspondence with anatomical connectivity. Little experimental evidence is available to answer this question. The connectivity graphs inferred by various statistical methods are commonly reported without examining their relation to anatomy. Topological properties of such graphs have been interpreted as principles of circuit organization (*e.g.* small-world organization) [62–64, 70]. However, the topological properties of functional connectivity graphs can depend on the method of inference [73]. Until a physiological interpretation of functional connectivity is established, the physiological relevance of such analyses remains in question and we did not attempt them here.

Inference of the conditional dependencies also depends on the completeness of the recorded population:

To equate conditional dependencies to direct interaction between neurons, we must record from all neurons with which the pair interacts. Unobserved portions of the circuit may manifest as conditional dependencies between observed neurons that do not interact. For this reason, statistical models of population activity have been most successfully applied to *in vitro* preparations of the retina or cell cultures where high-quality recordings from the complete populations were available [67]. In cortical tissue, electrode arrays record from a small fraction of cells in a given volume, limiting the validity of inference of the pairwise conditional dependencies. Perhaps for this reason, partial correlations have not, until now, been used to describe the functional connectivity in cortical populations.

Two-photon imaging of population calcium signals presents unique advantages for the estimation of functional connectivity. While the temporal resolution of calcium signals is limited by calcium dye kinetics, fast imaging techniques combined with spike inference algorithms provide millisecond-scale temporal resolution of single action potentials [74]. However, such high temporal precision comes at the cost of the accuracy of inferred spike rates. Better accuracy is achieved when calcium signals are analyzed on scales of tens of milliseconds [60]. 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* [59,60] and even from entire nervous systems [75,76]. These techniques may provide more incisive measurements of functional connectivity than electrophysiological recordings.

The low temporal resolution of calcium signals limits the use of functional connectivity methods that rely on millisecond-scale binning of signals (cross-correlograms, some GLMs, and binary maximum entropy models). Hence, most studies of functional connectivity have relied on instantaneous sample correlations [23,26,29,63]. Although some investigators have interpreted such correlations as indicators of (chemical or electrical) synaptic connectivity, most used them as more general indicators of functional connectivity without relating them to underlying mechanisms.

In this study, we sought to infer pairwise functional connectivity networks in cortical microcircuits. We hypothesized that partial correlations correspond more closely to underlying mechanisms than sample correlations when recordings are sufficiently dense. Since neurons form synaptic connections mostly locally and sparsely [77], we *a priori* favored solutions with sparse partial correlations. Under the assumptions that the recorded population is sufficiently complete and that the model correctly represents the nature of interactions, the network of partial correlations can be hypothesized to be a better representation of functional dependencies than correlations.

Functional connectivity as coactivations

Another approach to describing the functional connectivity of a circuit is to isolate individual patterns of multineuronal coactivations [78–83]. 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 [78,81], *i.e.* groups of tightly interconnected groups of cells involved in a common computation. Coactivation patterns could also result from shared input from unobserved parts of the circuit, or global network fluctuations modulating the activity of the local circuit [84].

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

Combining pairwise interactions and coactivations

In the effort to account for the joint activity patterns that are poorly explained by pairwise interactions, investigators have augmented models of pairwise interactions with additional mechanisms such as latent

variables [85], high-order correlations [86], or global network fluctuations [87].

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 [52, 53]. The resulting estimator, $C_{\text{sparse+latent}}$, effectively decomposed the functional connectivity into a sparse network of pairwise interactions and coactivation mode vectors.

Addressing ill-posedness

Inferring the conditional dependencies between variables in a probabilistic model is an ill-posed problem: small variations in the data produce large errors in the inferred network of dependencies. The problem becomes worse as the number of recorded neurons increases until such models lose their statistical validity [88]. As techniques have improved to allow recording from larger neuronal populations, experimental neuroscientists have addressed this problem by extending the recording durations to keep sampling noise in check and verified that existing models are not overfitted [87]. However, ambitious projects, such as the BRAIN initiative [48], aim to record from significantly larger populations. Simply increasing recording duration will not be sufficient, and the problem must be addressed by using regularized estimators. Regularization biases the solution toward a small subspace in order to counteract the effect of sampling noise in the empirical data. However, biasing the solution to an inappropriate subspace does not allow significant estimation improvement and hinders interpretation.

Several strategies have been developed to limit the model space in order to improve the quality of the estimate. For example, Ganmor et al. [86] developed a heuristic rule to identify the most significant features that must be fitted by a maximum entropy model for improved performance in the retina. As another example of regularization, generalized linear models typically employ L_1 penalty terms to constrain the solution space and to effectively reduce the dimensionality of the solution [67].

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

Model selection

Various model selection criteria have been devised to select between families of models and the optimal subsets of variables in a given model family based on observed data. Despite its computational requirements, cross-validation is among the most popular model selection methods due to its minimal assumptions about the data generating process [89].

We evaluated the covariance matrix estimators a loss function based on multivariate normal log likelihood. However, this does not limit the applicability of its conclusions to normal distributions. Indeed our major findings could be reproduced using other loss functions (compare Fig. 3 and Fig. S2). Other probabilistic models, fitted to the same data, could also serve as estimators of the covariance matrix. If a different model yields better estimation of the covariance matrix than the estimator proposed here, we believe that its structure should deserve consideration as the better representation of the functional connectivity.

The results of model selection must be interpreted with caution. As we demonstrated by simulation, even models with incorrect forms of dependencies can substantially improve estimates (Fig. 1). Therefore, showing that a more constrained model has better cross-validated performance than a more complex model does not necessarily support the conclusion that it reveals a better representation of dependencies in the data. This caveat is related to *Stein's Paradox* [90]: The biasing of an estimate toward an arbitrary low-dimensional target can consistently outperform a less constrained estimate.

Physiological interpretation and future directions

We showed that among several models a sparse network of linear interactions with several latent inputs yielded the best estimates of the noise covariance matrix for cortical microcircuits. This finding is valuable in itself: improved estimates of the noise covariance matrix for large datasets are important in order to understand the role of noise correlations in population coding [1, 6, 7, 11]

Moreover, this estimation approach provides a graphical representation of the dependencies in the data that can be used to formulate and test hypotheses about the structure of connectivity in the microcircuit. Importantly, the inferred functional interactions were substantially different from the network of the highest sample correlations. For example, the $C_{\text{sparse+latent}}$ estimator reveals a large number of negative interactions that were not present in the sample correlation matrix (Fig. 4 F) and may reflect inhibitory circuitry.

Distances between cells in physical space and in sensory feature space had a stronger effect on the partial correlations estimated by the $C_{\text{sparse+latent}}$ estimator than on sample correlations (Fig. 6 A–C). These differences support the idea that correlations are built up from partial correlations in chains of intermediate cells positioned closer and tuned more similarly to one another, with potentially closer correspondence to anatomical connectivity. These differences may also be at least partially explained by a trivial effect of regularization: the L_1 penalty applied by the estimator (Eq. 17) suppresses small partial correlations to greater extent than large partial correlations, enhancing the apparent effect of distance and tuning. Distances between cells in physical space and in sensory feature space also had distinct effects on the positive and negative connectivities inferred by the estimator (Fig. 6 D–F); we could not think of a numerical artifact that could produce this effect. The distinct positive and negative connectivity patterns may reflect geometric and graphical features of local excitatory and inhibitory networks. Indeed, the relationships between patterns of positive and negative connectivities inferred by the estimator mirrored the properties of excitatory and inhibitory synaptic connectivities with respect to distance, cortical layers, and feature tuning [23, 77, 91–96]. For example, while excitatory neurons form synapses within highly specific local cliques [77], inhibitory interneurons form synapses with nearly all excitatory cells within local microcircuits [23, 94, 97]. To further investigate the link between synaptic connectivity and inferred functional connectivity, in future experiments, we will use molecular markers for various cell types with follow-up multiple whole-cell *in vitro* recordings [23, 28] to directly compare the inferred functional connectivity graphs to the underlying anatomical circuitry. Finally, the latent units inferred by the estimator can be analyzed for their physiological functions. For example, these latent units may be modulated under different brain states (e.g. slow-wave sleep, attention) and stimulus conditions (e.g. certain types of stimuli may engage feedback connections).

Materials and Methods

Ethics statement

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

Surgery and two-photon imaging

The surgical procedures and data acquisition were performed as described in [60]: C57BL/6J mice (aged p40–60) were used. Anesthesia was initiated with isoflurane (3%) and maintained with 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 cortex. Membrane-permeant calcium indicator Oregon Green 488 BAPTA-1 AM (OGB-1, Invitrogen) was loaded by bolus injection. The craniotomy was sealed using a glass coverslip secured with dental cement.

Calcium imaging began 1 hour after dye injection. All imaging was performed using 3D-RAMP two-photon microscopy [60]. First, a 3D stack was acquired and cells were manually segmented. Then calcium signal were collected by sampling in the center of each cell at rates of 100 Hz or higher, depending on the number of cells.

Visual stimulus

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

Data processing

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

The measured fluorescent traces were deconvolved to reconstruct the firing rates for each neuron: First, the first principal component was subtracted from the raw traces in order to reduce common mode noise related to small cardiovascular movements [60]. The resulting traces were high-pass filtered above 0.1 Hz and downsampled to 20 Hz (Fig. 2 C). Then, the firing rates were estimated using by nonnegative deconvolution [61].

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

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

Cross-validation

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

Since each of the regularized estimators had one or two hyperparameters, we used *nested cross-validation*: The outer loop evaluated the performance of the estimators with the hyperparameter values optimized by cross-validation within the inner loop. Hyperparameters were optimized by a two-phase search algorithm: random search to find a good starting point for the subsequent pattern search to find the global minimum. The inner cross-validation loop subdivided the training dataset from the outer loop to perform 10-fold cross-validation in order to evaluate each choice of the hyperparameter values. Thus the size of the training dataset within the inner loop comprised 81% of the entire recording. Fig. S1

illustrates the dependence of the validation loss on the hyperparameters of the $C_{\text{sparse+latent}}$ estimator for the example site shown in Figures 2 and 4 and the optimal value found by the pattern search algorithm.

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

Covariance estimation

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

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

$$D = (1 - \alpha)C_{\text{sample}} \circ I + \alpha \frac{1}{p} \text{tr}(C_{\text{sample}})I \quad (12)$$

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

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

$$(L, D) = \arg \min_{L, \hat{D}} \mathcal{L}(\hat{L} + \hat{D}, C_{\text{sample}}), \quad (13)$$

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

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

$$S = \arg \min_{\hat{S} \succ 0} \mathcal{L}(\hat{S}^{-1}, C_{\text{sample}}) + \lambda \|\hat{S}\|_1 \quad (14)$$

where $\hat{S} \succ 0$ denotes the constraint that \hat{S} be a positive-definite matrix and $\|\hat{S}\|_1$ is the element-wise L_1 norm of the matrix \hat{S} . This problem formulation is known as *graphical lasso* [98, 99]. To solve this minimization problem, we adapted the alternative-direction method of multipliers (ADMM) [53]. Unlike C_{diag} and C_{factor} , this estimator does not include linear shrinkage: the selection of the sparsity level provides sufficient flexibility to fine-tune the regularization level.

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

$$S^* = \begin{pmatrix} S & S_{12} \\ S_{12}^T & S_{22} \end{pmatrix}, \quad (15)$$

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

$$C_{\text{sparse+latent}} = (S - S_{12}S_{22}^{-1}S_{12}^T)^{-1} \quad (16)$$

The rank of the $p \times p$ matrix $L = S_{12}S_{22}^{-1}S_{12}^T$ matches the number of the latent units in the joint distribution. Rather than finding S_{12} and S_{22} separately, L can be estimated as a low-rank matrix. We adapted the

ADMM algorithm to minimize the loss function with L_1 penalty on S to regulate its sparseness and nuclear norm penalty on L to regulate its rank [52, 53]:

$$(S, L) = \arg \min_{\hat{S}, \hat{L}} \mathcal{L}(\hat{S} - \hat{L}, C_{\text{sample}}) + \alpha \|\hat{S}\|_1 + \beta \text{tr}(\hat{L}) \quad (17)$$

The partial correlation matrix (Eq. 4) computed from $C_{\text{sparse+latent}}$ includes interactions between the visible and latent units and was used in Fig. 4 C and D and Fig. 5 C, and Fig. 6 A–C). The partial correlation matrix computed from S alone expresses strengths of pairwise interactions

$$P_{\text{sparse}} = -(S \circ I)^{-\frac{1}{2}} S (S \circ I)^{-\frac{1}{2}} \quad (18)$$

and were used in Fig. 4 F, G, H.

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

Cross-validation with conditioned variances

This study required an unbiased estimate of the discrepancy between various correlation matrix estimates and the true value of the correlation matrix. This was accomplished by cross-validation whereby regularized covariance matrix estimates computed from training samples were compared to unbiased covariance matrix estimates computed from independent testing samples. Loss functions were defined with respect to the covariance matrix rather than the correlation matrix because no unbiased estimate of the correlation matrix exists for finite sample sizes in the general case [100]. In particular, we showed that with the normal loss function (Eq. 9), the validation loss $\mathcal{L}(C, C'_{\text{sample}})$ is an unbiased estimate of true loss when C'_{sample} is an unbiased estimate. However, this approach only worked under the assumption of uniform variances across all conditions included in the training and testing samples. Here, we extend the validation loss to the estimation and evaluation of a common correlation matrix across multiple conditions with different variances.

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

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

$$C_{c,\text{sample}} = \frac{1}{n_c} \sum_{t \in T_c} (x(t) - \bar{x}_c) (x(t) - \bar{x}_c)^\top \quad (19)$$

and

$$C'_{c,\text{sample}} = \frac{1}{n'_c} \sum_{t \in T'_c} (x(t) - \bar{x}_c) (x(t) - \bar{x}_c)^\top \quad (20)$$

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

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

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

$$R_{\text{sample}} = \frac{1}{n} \sum_c n_c \left(V_{c,\text{sample}}^{-\frac{1}{2}} C_{c,\text{sample}} V_{c,\text{sample}}^{-\frac{1}{2}} \right) = \frac{1}{n} \sum_c \sum_{t \in T_c} z(t) z(t)^\top, \quad (21)$$

where $n = \sum_c n_c$ and $V_{c,\text{sample}} = C_{c,\text{sample}} \circ I$ is the diagonal matrix containing the sample variances. Then

R_{sample} is simply the covariance matrix of the z -score signal $z(t) = V_{c,\text{sample}}^{-\frac{1}{2}} (x(t) - \bar{x}_c)$ of the training sample.

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

$$C_{\text{sample}} = V_{\text{sample}}^{\frac{1}{2}} R_{\text{sample}} V_{\text{sample}}^{\frac{1}{2}} \quad (22)$$

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

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

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

$$C_{c,\text{reg}} = Q_c^{\frac{1}{2}} C_{\text{reg}} Q_c^{\frac{1}{2}} \quad (23)$$

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

The overall loss function is obtained by averaging the loss functions across all the conditions:

$$\frac{1}{\sum_c n'_c} \sum_c n'_c \mathcal{L}(C_{c,\text{reg}}, C'_{c,\text{sample}}) \quad (24)$$

With the normal loss $\mathcal{L}(\cdot, \cdot)$ (Eq. 9) and the unbiased testing covariance matrix $C_{c,\text{sample}}$, the loss function in Eq. 24 is an unbiased estimate of the true loss. Hence, it was used for evaluations reported in Fig. 3.

Simulation

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

Acknowledgments

We thank Genevra Allen for a helpful discussion, and Eftychios Pnevmatikakis for helpful suggestions and feedback on the manuscript.

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. Ko H, Cossell L, Baragli C, Antolik J, Clopath C, et al. (2013) The emergence of functional microcircuits in visual cortex. *Nature* 496: 96–100.
29. 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.
30. Goard M, Dan Y (2009) Basal forebrain activation enhances cortical coding of natural scenes. *Nat Neurosci* 12: 1444–9.
31. Kohn A, Zandvakili A, Smith MA (2009) Correlations and brain states: from electrophysiology to functional imaging. *Curr Opin Neurobiol* 19: 434–8.
32. 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.
33. Perkel DH, Gerstein GL, Moore GP (1967) Neuronal spike trains and stochastic point processes: II. simultaneous spike trains. *Biophysical journal* 7: 419–440.
34. Moore GP, Segundo JP, Perkel DH, Levitan H (1970) Statistical signs of synaptic interaction in neurons. *Biophysical Journal* 10: 876–900.
35. Shadlen MN, Newsome WT (1998) The variable discharge of cortical neurons: implications for connectivity, computation, and information coding. *J Neurosci* 18: 3870–96.
36. Salinas E, Sejnowski TJ (2001) Correlated neuronal activity and the flow of neural information. *Nature Reviews Neuroscience* 2: 539–550.
37. 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.
38. Rosenbaum R, Josić K (2011) Mechanisms that modulate the transfer of spiking correlations. *Neural computation* 23: 1261–1305.
39. Whittaker J (1990) Graphical models in applied multivariate statistics. Wiley Publishing.
40. 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.

41. Peng J, Wang P, Zhou N, Zhu J (2009) Partial correlation estimation by joint sparse regression models. *Journal of the American Statistical Association* 104.
42. 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.
43. 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.
44. Ledoit O, Wolf M (2004) A well-conditioned estimator for large-dimensional covariance matrices. *Journal of multivariate analysis* 88: 365–411.
45. Bickel PJ, Li B, Tsybakov AB, van de Geer SA, Yu B, et al. (2006) Regularization in statistics. *Test* 15: 271–344.
46. 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.
47. Friedman JH (1989) Regularized discriminant analysis. *Journal of the American statistical association* 84: 165–175.
48. Alivisatos AP, Chun M, Church GM, Deisseroth K, Donoghue JP, et al. (2013) The brain activity map. *Science* 339: 1284–1285.
49. Anderson T (2003) An introduction to multivariate statistical analysis. *Wiley series in probability and statistics* .
50. Dempster A (1972) Covariance selection. *Biometrics* : 157–175.
51. Baba K, Shibata R, Sibuya M (2004) Partial correlation and conditional correlation as measures of conditional independence. *Australian & New Zealand Journal of Statistics* 46: 657–664.
52. 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.
53. Ma S, Xue L, Zou H (2013) Alternating direction methods for latent variable gaussian graphical model selection. *Neural computation* : 1–27.
54. 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.
55. Fan J, Fan Y, Lv J (2008) High dimensional covariance matrix estimation using a factor model. *Journal of Econometrics* 147: 186–197.
56. Vogels R, Spileers W, Orban G (1989) The response variability of striate cortical neurons in the behaving monkey. *Experimental brain research* 77: 432–436.
57. Ponce-Alvarez A, Thiele A, Albright TD, Stoner GR, Deco G (2013) Stimulus-dependent variability and noise correlations in cortical mt neurons. *Proceedings of the National Academy of Sciences* 110: 13162–13167.
58. Reddy GD, Saggau P (2005) Fast three-dimensional laser scanning scheme using acousto-optic deflectors. *J Biomed Opt* 10: 064038.

59. 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* .
60. Cotton RJ, Froudarakis E, Storer P, Saggau P, Tolias AS (2013) Three-dimensional mapping of microcircuit correlation structure. *Frontiers in Neural Circuits* 7: 151.
61. 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.
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. 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.
64. Sadovskiy AJ, MacLean JN (2014) Mouse visual neocortex supports multiple stereotyped patterns of microcircuit activity. *The Journal of Neuroscience* 34: 7769–7777.
65. Alonso JM, Martinez LM (1998) Functional connectivity between simple cells and complex cells in cat striate cortex. *Nature neuroscience* 1: 395–403.
66. Aertsen A, Gerstein G, Habib M, Palm G (1989) Dynamics of neuronal firing correlation: modulation of “effective connectivity”. *Journal of neurophysiology* 61: 900–917.
67. 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.
68. 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.
69. Tkacik G, Schneidman E, Berry I, Michael J, Bialek W (2006) Ising models for networks of real neurons. *arXiv preprint q-bio/0611072* .
70. Yu S, Huang D, Singer W, Nikolic D (2008) A small world of neuronal synchrony. *Cereb Cortex* 18: 2891–901.
71. 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.
72. 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.
73. Zalesky A, Fornito A, Bullmore E (2012) On the use of correlation as a measure of network connectivity. *Neuroimage* 60: 2096–2106.
74. 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.
75. Leung LC, Wang GX, Mourrain P (2013) Imaging zebrafish neural circuitry from whole brain to synapse. *Frontiers in neural circuits* 7.
76. 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.

77. 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.
78. Gerstein GL, Bedenbaugh P, Aertsen AM (1989) Neuronal assemblies. *Biomedical Engineering, IEEE Transactions on* 36: 4–14.
79. Chapin JK, Nicolelis MA (1999) Principal component analysis of neuronal ensemble activity reveals multidimensional somatosensory representations. *Journal of neuroscience methods* 94: 121–140.
80. 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.
81. 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.
82. Lopes-dos Santos V, Conde-Ocazonez S, Nicolelis MA, Ribeiro ST, Tort AB (2011) Neuronal assembly detection and cell membership specification by principal component analysis. *PloS one* 6: e20996.
83. Lopes-dos Santos V, Ribeiro S, Tort AB (2013) Detecting cell assemblies in large neuronal populations. *Journal of Neuroscience Methods* 220: 149 – 166.
84. 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.
85. 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* .
86. 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.
87. 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* .
88. 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.
89. Arlot S, Celisse A (2010) A survey of cross-validation procedures for model selection. *Statistics Surveys* 4: 40–79.
90. Efron B, Morris CN (1977) Stein's paradox in statistics. *WH Freeman*.
91. Song S, Sjöström PJ, Reigl M, Nelson S, Chklovskii DB (2005) Highly nonrandom features of synaptic connectivity in local cortical circuits. *PLoS biology* 3: e68.
92. Oswald AMM, Reyes AD (2008) Maturation of intrinsic and synaptic properties of layer 2/3 pyramidal neurons in mouse auditory cortex. *Journal of neurophysiology* 99: 2998.
93. Adesnik H, Scanziani M (2010) Lateral competition for cortical space by layer-specific horizontal circuits. *Nature* 464: 1155–1160.
94. Fino E, Yuste R (2011) Dense inhibitory connectivity in neocortex. *Neuron* 69: 1188-203.
95. Isaacson JS, Scanziani M (2011) How inhibition shapes cortical activity. *Neuron* 72: 231–243.

96. Levy RB, Reyes AD (2012) Spatial profile of excitatory and inhibitory synaptic connectivity in mouse primary auditory cortex. *J Neurosci* 32: 5609-19.
97. Packer AM, Yuste R (2011) Dense, unspecific connectivity of neocortical parvalbumin-positive interneurons: a canonical microcircuit for inhibition? *The Journal of Neuroscience* 31: 13260–13271.
98. Meinshausen N, Bühlmann P (2006) High-dimensional graphs and variable selection with the lasso. *The Annals of Statistics* 34: 1436–1462.
99. Friedman J, Hastie T, Tibshirani R (2008) Sparse inverse covariance estimation with the graphical lasso. *Biostatistics* 9: 432–441.
100. Fisher RA (1921) On the “probable error” of a coefficient of correlation deduced from a small sample. *Metron* 1: 3–32.

Figure Legends

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

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

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

Row 4: Estimates computed from the same data as in Row 3 using structured estimators of the correct type, optimized by cross-validation. The regularized estimates are closer to the truth than the sample correlation matrices.

Row 5: Excess losses (Eq. 9) for the five estimators as a function of sample size. The error bars indicate the standard error of the mean based on 30 samples. Estimators with structure that matches the true model converged to zero faster than the other estimators.

Row 6: Validation losses for the five estimators relative to the matching estimator. Error bars indicate the standard error of the mean based on 30 samples. Differences in validation loss approximate differences in true loss.

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–F**. Calcium activity data from an example site. **C**. Representative calcium signals of seven cells, downsampled to 20 Hz, out of the 292 total recorded cells. Spiking activity inferred by nonnegative deconvolution is shown by red ticks below the trace. **D**. The spatial arrangement and orientation tuning of the 292 cells from the imaged site. The cells’ colors indicate their orientation preferences. The gray cells were not significantly tuned. **E**. The sample noise correlation matrix of the activity of the neural population. **F**. Histogram of noise correlation coefficients in one site. The red line indicates the mean correlation coefficient of 0.020.

Figure 3. Performance of estimator $C_{\text{sparse+latent}}$ with respect to the normal loss function (eq. 9) relative to the other estimators: C_{sample} , C_{diag} , C_{factor} , and C_{sparse} . Covariance estimators C_{sample} , C_{diag} , C_{factor} , and C_{sparse} produced consistently greater validation losses than $C_{\text{sparse+latent}}$ ($p < 0.01$ in each comparison, Wilcoxon signed rank test, $n = 27$ sites in 14 mice). The box plots indicate the 25th, 50th, and 75th percentiles with the whiskers extending to the minimum and maximum values after excluding the outliers marked with ‘+’.

Figure 4. Structure revealed by $C_{\text{sparse+latent}}$. **A, B**. The regularized estimate $C_{\text{sparse+latent}}$ closely approximates the sample correlation matrix C_{sample} . **C, D**. The partial correlation matrices from the two estimates differ substantially. **E**. The partial correlation matrix of the regularized estimate is decomposed into a sparse component with 92.8% off-diagonal zeros (bottom-left) and low-rank component of rank 72 (top-right). **F**. The sparse component of the regularized partial correlation matrix had little resemblance to the sample correlations: The gray region indicates the range of correlations containing 92.8% of cells pairs, equal to the fraction of zeros in the sparse partial correlation matrix. Correlation coefficients outside this interval formed the network of greatest correlations. This network differed from the sparse component of the $C_{\text{sparse+latent}}$: Only 27.7% of the highest correlations coefficients outside the gray regions coincided with interactions inferred by $C_{\text{sparse+latent}}$. **G**. A graphical depiction of the positive (green) and negative (magenta) sparse partial correlations as edges between observed neurons. The line density is proportional to the magnitude of the partial correlation. **H**. A subset of neurons from the center of the cluster shown in **G** showing the sparse partial correlations. **I**. The same subset of neurons with edges indicating sample correlations thresholded to match the sparsity of the sparse partial correlation. These edges correspond to the sample correlation coefficients outside the gray region in panel F.

Figure 5. Properties of $C_{\text{sparse+latent}}$ estimates from all imaged sites. Each point represents an imaged site with its color indicating the population size as shown in panels A and B. The example site from Figures 2 and 4 is circled in blue.

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

Figure 6. Dependence of sample correlations, regularized partial correlations, and connectivity inferred by $C_{\text{sparse+latent}}$ on the differences in preferred orientations, Δori , and physical distances, Δx and Δz . The error bars mark the standard errors of the means. Five sites with highest connectivity (see Fig. 5 B) were selected for this analysis. **A–C.** Normalized mean sample correlations (black) and normalized mean partial correlations (red) estimated by $C_{\text{sparse+latent}}$ across $n = 5$ imaged sites. The values in each bin are normalized by the means across the entire site, shown in Fig. 5 C, to make the effects more comparable across the sites. **A.** Mean partial correlations in $C_{\text{sparse+latent}}$ depend on Δori more strongly than mean sample correlations. **B.** Mean partial correlations in $C_{\text{sparse+latent}}$ depend on Δx more strongly than mean sample correlations. Only horizontally aligned cell pairs with $\Delta z < 30 \mu m$ were considered. **C.** Mean partial correlations in $C_{\text{sparse+latent}}$ depend on Δz more strongly than mean sample correlations. Only vertically aligned cell pairs with $\Delta x < 30 \mu m$ were considered. **D–F.** Normalized positive connectivity (green) and normalized negative connectivity (dark red) inferred by the $C_{\text{sparse+latent}}$ estimator in $n = 5$ imaged sites. Here “connectivity” refers to the fraction of the non-zero elements in the sparse component S of the $C_{\text{sparse+latent}}$ estimator, which describes inferred direct interactions between specific pairs of the recorded neurons. Positive and negative connectivities refer to the fractions of the positive and negative partial correlations computed from S , respectively. “Normalized positive connectivity” is the ratio of the positive connectivity for pairs meeting a given condition, *e.g.* similar tuning with $\Delta ori < 15^\circ$, to the average connectivity over the entire site. Normalized negative connectivity is computed similarly for the negative connectivity. The average connectivity across sites is shown in Fig. 5 B with only the five most connected sites included in the analysis. The normalization made the effects of tuning and distance more comparable across sites. **D.** Positive connectivity decreases with Δori whereas negative connectivity does not. **E.** Positive connectivity decays with Δx whereas negative connectivity does not, within the examined range. **F.** Positive connectivity decays with Δz . Negative connectivity does not decay for small values of Δz .

Supporting Information Legends

Figure S1. Optimization of hyperparameters of the $C_{\text{sparse+latent}}$ estimator. **A.** Validation loss (Eq. 24) for the example site in Fig. 2 and 4 as a function of the hyperparameters α and β of the $C_{\text{sparse+latent}}$ estimator (Eq. 8 and Eq. 17). In all panels, the red cross marks the optimal value found by the pattern search algorithm described in Methods. **B.** The connectivity ($1 - \text{sparsity}$) of the sparse component S as a function of α and β for the example site. **C.** The number of latent units, *i.e.* the rank of the low-rank component L , as a function of hyperparameters α and β . **D.** The loss function as a function of the connectivity and the number of latent units.

Figure S2. Performance of estimator $C_{\text{sparse+latent}}$ with respect to quadratic loss (Eq. 11) relative to the other estimators: C_{sample} , C_{diag} , C_{factor} , and C_{sparse} . All comparisons showed significant advantage of estimator $C_{\text{sparse+latent}}$ ($p < 10^{-3}$ in each comparison, Wilcoxon signed rank test, $n = 27$ sites in 14 mice).