

Improved estimation and interpretation of correlations in neural circuits
 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: atolias@cnsbcm.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 dense recordings, the correlation matrix should be best represented as the combination of a sparse matrix of pairwise partial correlations representing local interactions and a low-rank component representing common fluctuations and external inputs. Indeed, in cross-validation tests, the covariance matrix estimator with this structure consistently outperformed other regularized estimators. The sparse component of the estimate expressed a graph of interactions organized with respect to 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 the 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 dense recordings than the sample correlation matrix.

Author Summary

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

of large and dense groups of neurons in mouse visual cortex, the best estimation performance was attained by decomposing the correlation matrix into a sparse network of partial correlations ('interactions') combined with a low-rank component. The inferred interactions were both positive ('excitatory') and negative ('inhibitory') and reflected the spatial organization and orientation preferences of the interacting cells. We propose that the 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 with no simple mechanistic interpretation, it can be used to generate hypotheses about the anatomical architecture of the neural circuit and 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 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. When a large fraction of the interacting variables are observed, partial correlations more closely reflect the direct causal effects between components of complex systems than correlations. Partial correlations have been used to describe interactions between genes in functional genomics [39, 40] and between brain regions in imaging studies [41, 42]. These opportunities have not yet been explored in neurophysiological studies, as most studies 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) [43], and poorly conditioned partial correlations [39]. 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 [39, 44]. 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 [39, 43, 45] or by cross-validation [46]. 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 [45], functional genomics [39], and brain imaging [42], 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 [46]. But reduced estimation error is not the only benefit of regularization. The amount of improvement of estimation is greatest when the structure imposed by regularization limits the solution space, but still represents the dominant dependencies in the data. Therefore, by finding the most efficient among many regularized estimators, we learn something about the system itself: the structure of the most efficient estimator is a parsimonious representation of the regularities in the data.

The necessity and benefits of regularization increase with the size of the recorded population. With the advent of big neural data [47], 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 that may involve combinations of heuristic rules and numerical techniques specially designed for each neural circuit.

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_0 – sample covariance matrix, the unbiased, unregularized estimator.

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

C_{factor} – linear shrinkage of the covariances toward a low-rank matrix, representing inputs from unobserved 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 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, which we call ‘interactions’ here, between the observed neurons; it also inferred several latent units exerting linear effects on the observed neurons. We analyzed these networks of partial correlations and found the following: Whereas 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

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

where $\mathbb{E} [\cdot]$ denotes expectation; the $p \times 1$ vector x is a single observation of the firing rates of p neurons over time Δt ; and μ is the vector of expected firing rates. The usual way to estimate the covariance matrix is to calculate the *sample covariance matrix* C_0 from the empirical sample of observations $x(t)$, $t = 1, \dots, n$ 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 in one neuron's activity within the sample ($\nu = n - 1$ if observations are independent). For noise correlations, the true mean μ and the sample mean \bar{x} are conditioned on the stimulus. 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 the sample correlation matrix, C_0 .

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

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

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

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

We considered four regularized estimators based on distinct families of reduced target estimates: C_{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_0 toward D controlled by the scalar *shrinkage intensity* parameter $\lambda \in [0, 1]$:

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

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

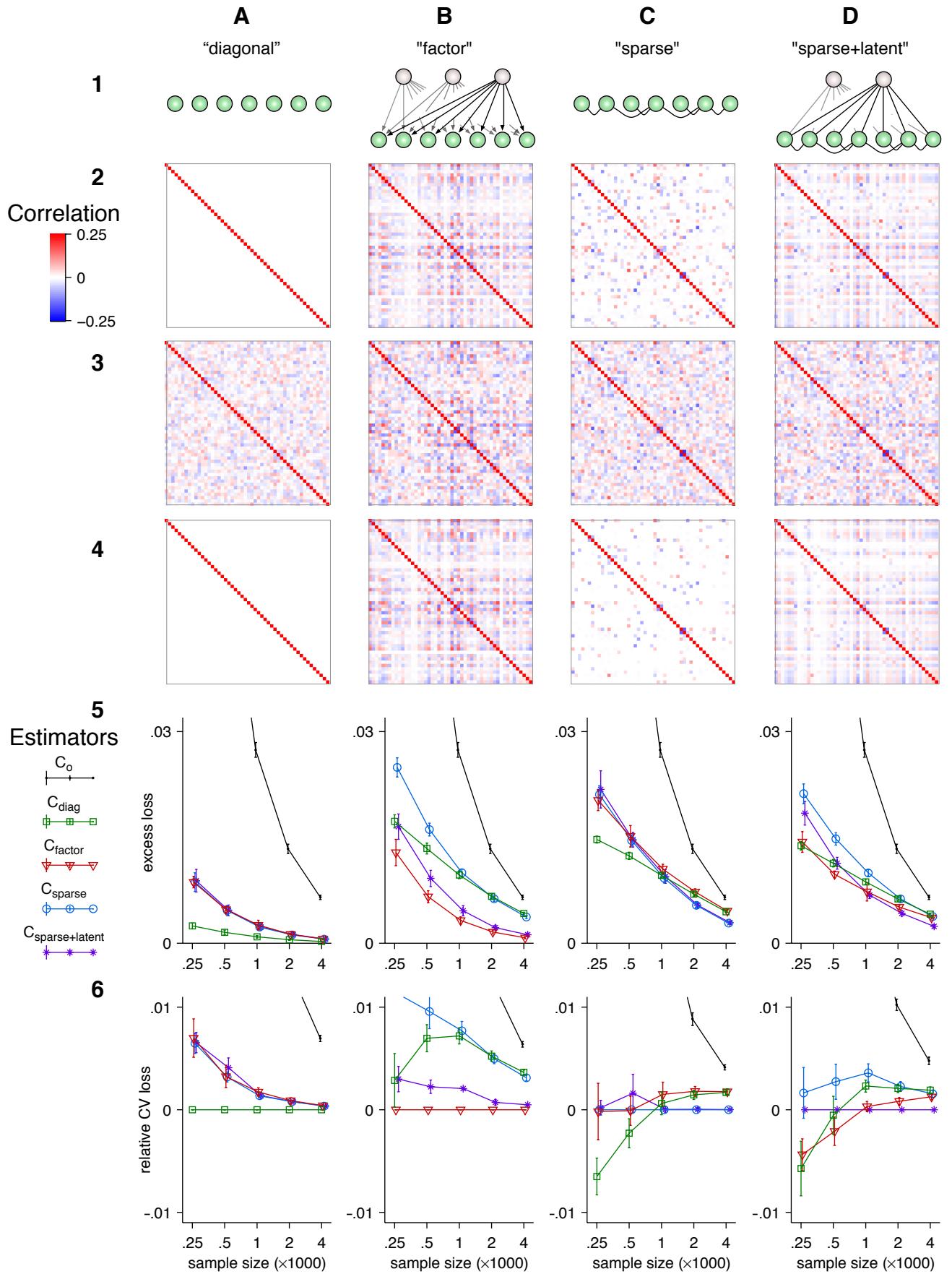
The target of estimator C_{factor} is the factor model $L + \Psi$, where L is the $p \times p$ positive definite matrix with low rank and Ψ is the diagonal matrix of independent variances. The estimator,

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

has two hyperparameters: the rank of L (*i.e.* the number of latent factors) and the shrinkage intensity λ . 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* [48]. 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 . 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) [49, 50].

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 50×50 covariance matrices, each with structure

Figure 1 (preceding page). Regularized estimators whose structure matches the true structure in the data are more efficient. **Row 1.** Graphical representations of the target estimates of the four respective regularized covariance matrix estimators. Recorded neurons are represented by green spheres and latent units by light-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* with 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 two latent units, **C.** a correlation matrix with 85% off-diagonal zeros in its inverse, and **D.** a correlation matrix whose inverse is the sum of a rank-1 (*i.e.* one latent unit) matrix and a sparse matrix with 89% off-diagonal zeros.

Row 3: Sample correlation matrices calculated from samples of size $n = 1000$ drawn from simulated random processes with respective correlation matrices shown in Row 2. The structure of the sample correlation matrix is difficult to discern.

Row 4: Correlation matrix estimates computed by estimators with matching structure from the same data as the sample correlation matrices in Row 3. The regularized estimates are closer to the truth than the sample correlation matrices.

Row 5: Excess losses (Eq. 11) 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. Some exceptions occur when the sample is too small to reveal the structure in the data. In these cases estimators with simpler structures can be more efficient than estimators with true structure.

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

that matched one of the four regularized estimators (Fig. 1 Row 2, A–D, see Methods for details about how these matrices were generated). We used these covariance matrices as the ground truth in multivariate normal distributions with zero means and drew samples of various sizes. Even with relatively large sample sizes (*e.g.* $n = 1000$), the sample correlation matrices were contaminated by sampling noise and their underlying 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 $L(\Sigma|C_0)$,

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

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

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

The choice of the normal loss function is motivated by mathematical convenience. We expect that the main conclusions of our study will not change qualitatively with other well behaved loss functions, such as Stein’s entropy loss or quadratic loss [39, 43, 51, 52].

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

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

We drew 30 independent samples of sizes $n = 250, 500, 1000, 2000$, and 4000 from each model and computed the excess loss 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 typically outperformed the other estimators (Fig. 1 row 5). There were two exceptions: First, for some smaller samples, the data were insufficient to reveal the true correlation structure and the estimators with simpler structures, C_{diag} and C_{factor} , often outperformed the estimators with the matching structure (Fig. 1 Row 5, C and D). Second, $C_{\text{sparse+latent}}$ often performed as well as C_{sparse} on data from models without latent units because it reliably inferred the correct number of latent units: zero.

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 must 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'_0 to validate the estimate C computed from the *training sample*. The estimate C'_0 is used to compute the *validation loss* $\mathcal{L}(C, C'_0)$, which approximates the loss $\mathcal{L}(C, \Sigma)$.

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

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

With this property of additivity, the validation loss is an unbiased estimate of the loss function:

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

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

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

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

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

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

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

The relative cross-validation loss was used for the main findings in this study.

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

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 $C_{sparse+latent}$ estimator is most efficient in neural data We recorded the calcium activity of dense 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) [53–55]. This technique allowed fast sampling (100–150 Hz) from large numbers (150–350) of cells in a small volume of cortical tissue ($200 \times 200 \times 100 \mu\text{m}^3$) in layers 2/3 and 4 (Fig. 2 C). The firing rates were inferred using sparse nonnegative deconvolution [56] (Fig. 2 C and D). Only cells that produced detectable calcium activity were included in the analysis (For details, see Methods). First, 30 repetitions of full-field drifting gratings of 16 directions were presented in random order. Each grating was presented for 500 ms, without intervening blanks. This stimulus was used to compute the orientation tuning of the recorded cells (Fig. 2 E). To estimate the noise correlation matrix, we presented only 2 distinct directions in some experiments or 5 directions in others with between 100 and 300 repetitions of each direction. In this stimulus, each grating lasted 1 second and was followed by a 1-second blank. The average stimulus response was subtracted from each trial. The traces were then binned into 150 ms intervals for the estimation of the covariance matrix. The sample correlation coefficients were largely positive and low (average across sites of 0.02 and under 0.04 in any site) (Fig. 2 G and Fig. 5 D).

In these dense 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 $C_{sparse+latent}$) 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 $C_{sparse+latent}$. If both types of effects are significant, then $C_{sparse+latent}$ should outperform the other estimators.

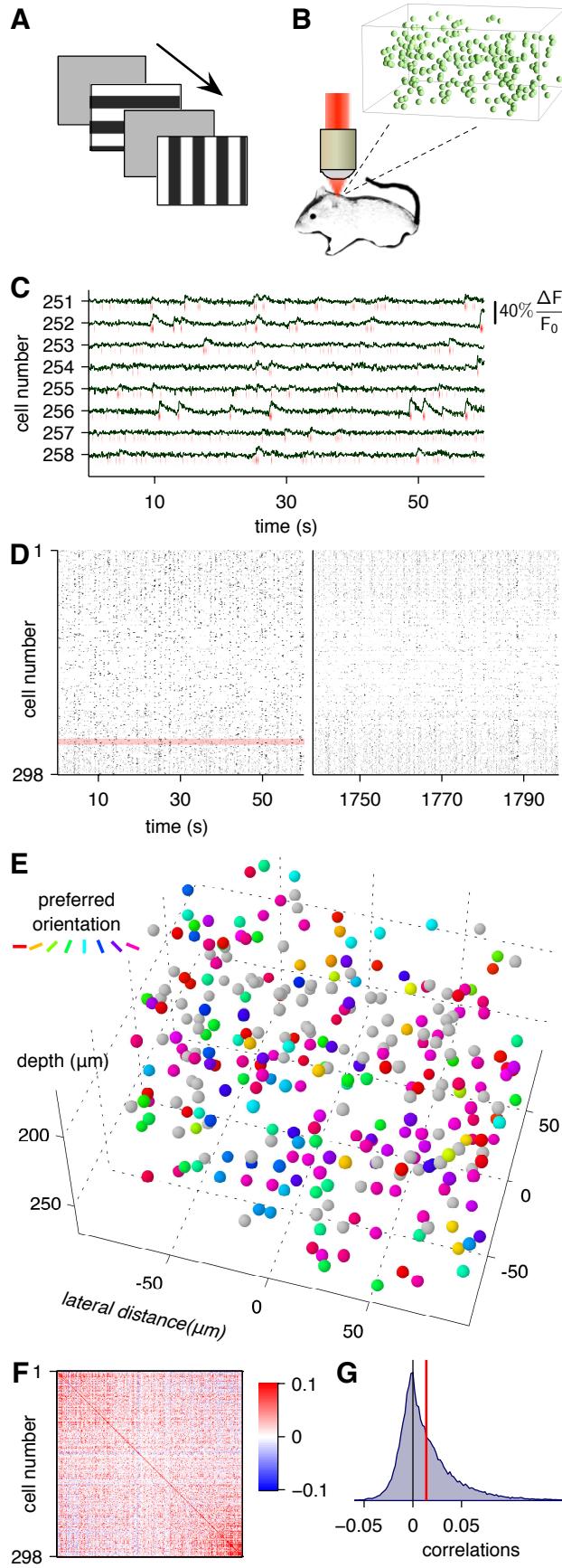


Figure 2. Acquisition of neural signals for the estimation of noise correlations. Visual stimuli comprising full-field drifting gratings interleaved with blank screens (A) presented during two-photon recordings of somatic calcium signals using fast 3D random-access microscopy (B). C–G. Calcium activity data from an example site. C. Representative calcium signals from eight cells out of 298 cells downsampled to 20 Hz. The inferred firing rate binned in 150 ms intervals are indicated by red ticks below each trace. D. The raster plot of the inferred firing rates, binned in 150 ms intervals, from the entire population from the first (left) and last (right) minute of the entire recording. The traces from C are highlighted in red. E. The spatial arrangement and orientation tuning of the 298 cells from the imaged site. Only spiking cells are shown. The cells' colors indicate their orientation preferences. The gray cells were not significantly tuned. F. The sample noise correlation matrix of the activity of the neural population. G. Histogram of noise correlation coefficients. The red line indicates the mean.

To test this hypothesis, we computed the relative CV losses (Eq. 13) of estimators C_0 , C_{diag} , C_{factor} , and C_{sparse} with respect to $C_{\text{sparse+latent}}$ in 31 imaged sites in 15 mice. The hyperparameters of each estimator were optimized by nested cross-validation (See Methods for details). Indeed, the median relative CV loss was significantly above zero (Fig. 3), indicating that $C_{\text{sparse+latent}}$ produced estimates that were consistently closest to the true covariance matrix.

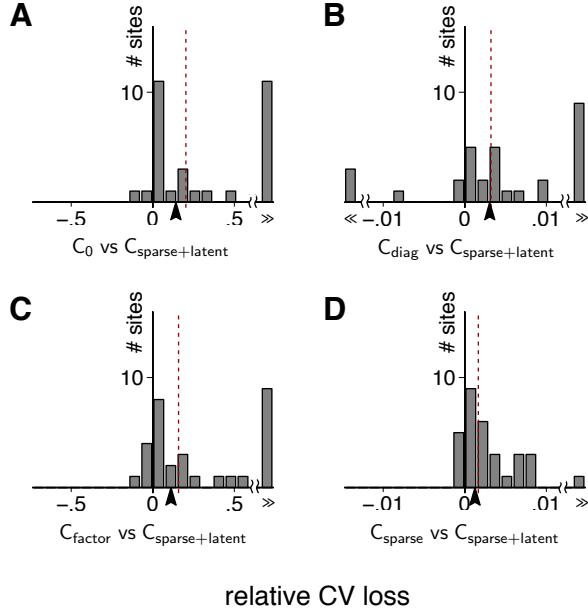


Figure 3. Estimator $C_{\text{sparse+latent}}$ is more efficient than the other estimators on neural data.

A–D. Histograms of the relative cross-validation loss of estimators C_0 , C_{diag} , C_{factor} , and C_{sparse} with respect to $C_{\text{sparse+latent}}$. The histograms were obtained from 31 imaged sites in 15 mice. All medians (red dashed lines) were significantly greater than zero, indicating the dominance of $C_{\text{sparse+latent}}$ over the other estimators. The arrow heads indicate the results for the site shown in Fig. 2 and Fig. 4.

A. $C_{\text{sparse+latent}}$ outperforms C_0 : median improvement 0.16 nats/neuron, $p = 3.9 \times 10^{-5}$.

B. $C_{\text{sparse+latent}}$ outperforms C_{diag} : median improvement 0.0032 nats/neuron, $p = 2.2 \times 10^{-3}$.

C. $C_{\text{sparse+latent}}$ outperforms C_{factor} : median improvement 0.16 nats/neuron, $p = 6.9 \times 10^{-5}$.

D. $C_{\text{sparse+latent}}$ outperforms C_{sparse} : median improvement 0.0016 nats/neuron, $p = 5.5 \times 10^{-6}$.

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

The same evaluation based on the quadratic loss function,

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

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 The estimator $C_{\text{sparse+latent}}$ separates two sources of correlations: a network of linear interactions between pairs of neurons and a set of latent units reflecting common fluctuations across the entire population.

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 D), the corresponding partial correlation matrices differed substantially (Fig. 4 B and E). The partial correlation matrix of the regularized estimate was the sum of a sparse component and a low-rank component (Fig. 4 C).

In the example site (Fig. 4), the sparse component had 82.2% sparsity (or conversely, 17.8% connectivity: $\text{connectivity} = 1 - \text{sparsity}$) with average node degree of 52.9 (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 15, denoting 15 inferred latent units. The number of latent units and the average node degrees generally increased with population size, although these values varied widely between the datasets (Fig. 5 A and B). There was an inverse relationship between the number of latent units and the average node degree (Fig. 5 C): 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 needs further investigation.

The average partial correlations calculated from these estimates according to Eq. 4 at all 31 sites were about 5 times lower than the average sample correlations (Fig. 5 D). This suggests that correlations between neurons build up from large numbers of small interactions with other neurons. Furthermore, the average partial correlations were less variable: the coefficient of variation of the average sample correlations across sites was 0.41 whereas that of the average partial correlations was 0.35.

While the sample correlations were mostly positive, the sparse component of the partial correlations ('interactions') had a high fraction (43.3% in the example site) of negative values (Fig. 4 F). This fraction was consistent across multiple sites but the fraction of negative interactions increased with the inferred connectivity Fig. 5 F. This increase suggests that negative interactions are inferred only after a sufficient density of positive interactions has been uncovered.

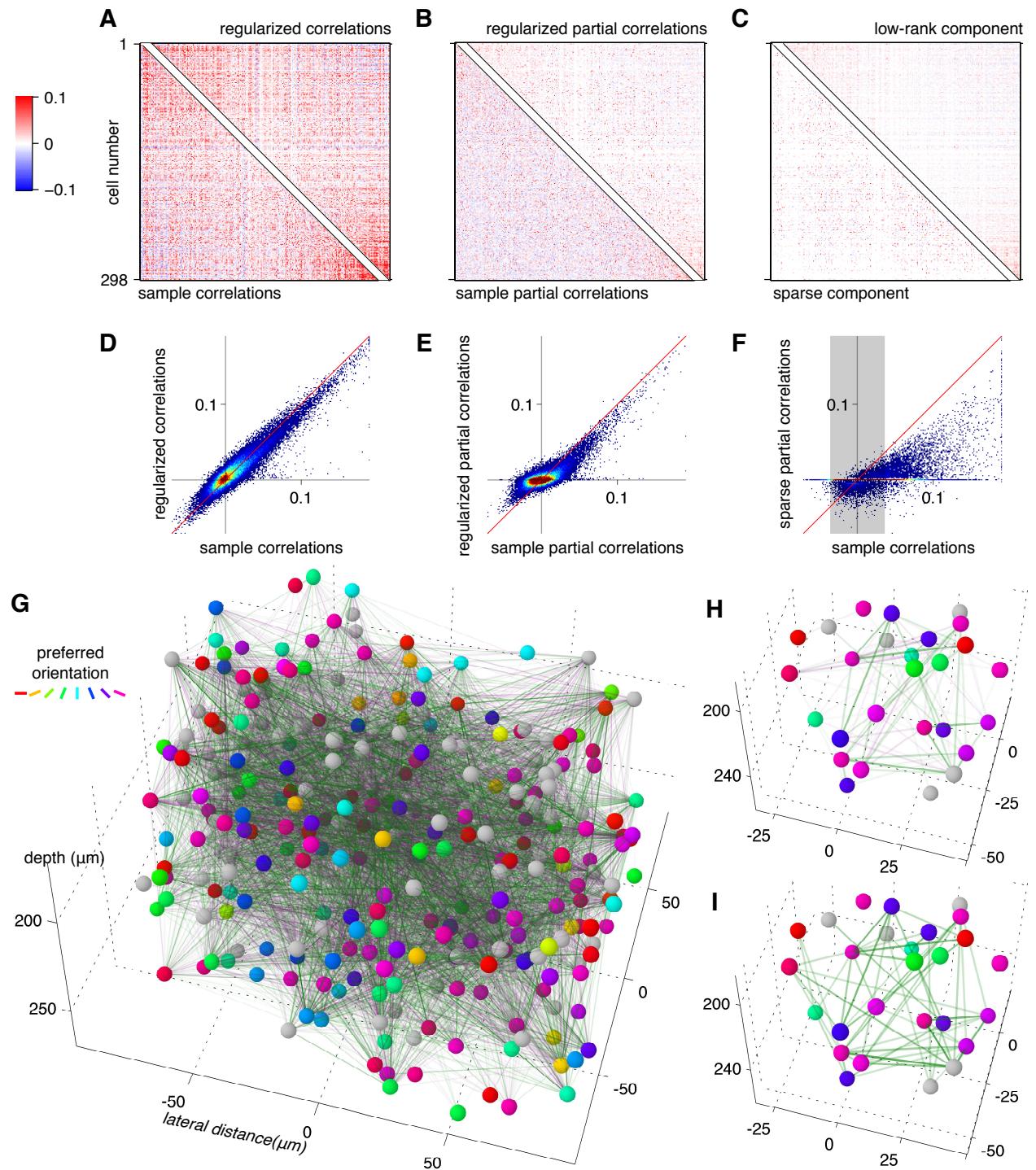
Thresholded sample correlations have been used in several studies to infer pairwise interactions [26, 57, 58]. We therefore compared the interactions inferred by $C_{\text{sparse+latent}}$ to those obtained by thresholding the sample correlations with the threshold set to match the connectivity of the $C_{\text{sparse+latent}}$ graph. The networks revealed by the two methods differed substantially. In the example site with 17.8% connectivity inferred by $C_{\text{sparse+latent}}$, only 41.7% of connections coincided with 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 were expected given the rest of the correlation matrix. This low overlap between the two types of graphs was consistent across the datasets (Fig. 5 E).

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

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

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 was 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, 59]. Such signatures are ambiguous as they can arise from network effects other than direct synaptic connections [60]. 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, conditionally independent neurons can be inferred to lack 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 and membrane kinetics, and individual neurons' stimulus drive [61]. Maximum entropy models constrained by observed pairwise correlations also model pairwise coupling between cells [62–66]. 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 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) [57, 58, 64]. However, the topological properties of functional connectivity graphs can depend on the method of inference [67]. Until a physiological interpretation of functional connectivity is established, the validity 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 properly condition the pairwise interactions on the activity of the other neurons, we must record from all neurons with which the pair interacts. Unobserved portions of the circuit can introduce false conditional dependencies between observed neurons. 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 [61]. 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. This may be the reason that partial correlations have not previously been used to describe the functional connectivity in cortical populations until this study.

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 [68]. 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 [55]. The major advantage of calcium imaging is its ability to characterize the

Figure 4 (preceding page). Example of the structure revealed by $C_{\text{sparse+latent}}$ the sparse+latent estimator. A and D. The regularized estimate $C_{\text{sparse+latent}}$ closely approximates the sample correlation matrix C_0 . **B and E.** However, the partial correlation matrices from the two estimates differ substantially. **C.** 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. The significant correlations were outside this interval. Yet 58.3% of the interactions inferred by $C_{\text{sparse+latent}}$ linked pairs of neurons whose correlation was below the threshold. Only the remaining 41.7% of the interactions overlapped with sample correlations above the threshold. **G.** A graphical depiction of the positive (green) and negative (magenta) partial correlations as edges between observed neurons. The line density is proportional to the magnitude of the correlation. **H.** A subset of neurons from the center of the cluster shown in **G** showing the regularized partial correlations. **I.** The same subset with sample correlations thresholded to match the sparsity of the regularized interactions.

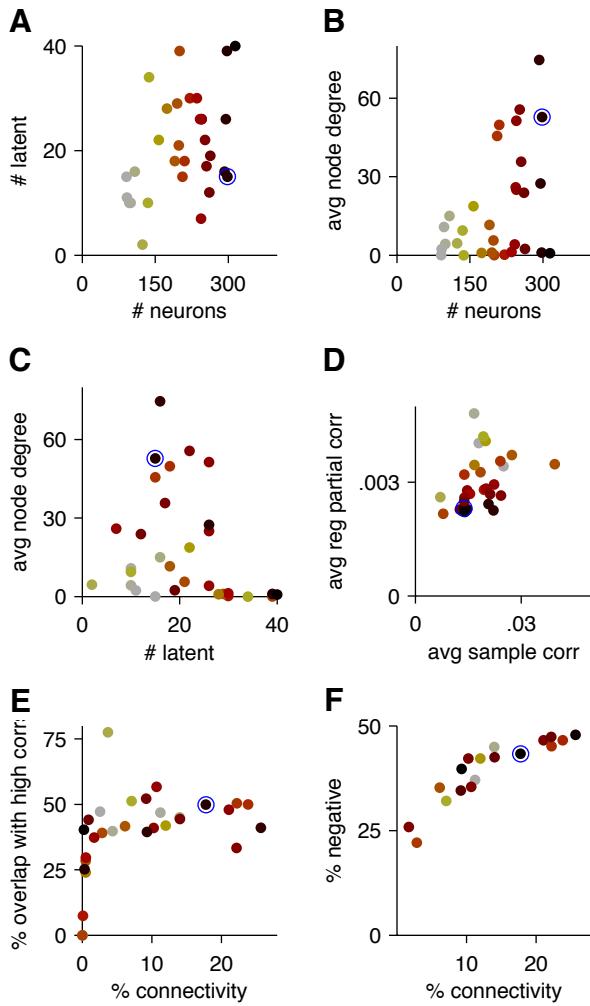


Figure 5. Properties of $C_{\text{sparse+latent}}$ estimates from all imaged sites. Each point represents an imaged site with its color indicating the population size as shown in panels A and B. The example site from Figures 2 and 4 is circled in blue.

A. The number of inferred latent units vs. population size. **B.** The average node degree for sparse partial correlations vs. population size. **C.** The average node degree in the sparse component in each site vs. the number of latent units described by the low-rank component of the estimate. **D.** The average sample correlations vs. the average partial correlations (Eq. 4) of the $C_{\text{sparse+latent}}$ estimate. **E.** Percent overlap between the graph of highest sample correlations and the graph of interactions inferred by $C_{\text{sparse+latent}}$ as a function of the connectivity (*i.e.* $1 - \text{sparsity}$) of its sparse component. **F.** The percentage of negative interactions vs. connectivity in the $C_{\text{sparse+latent}}$ estimates.

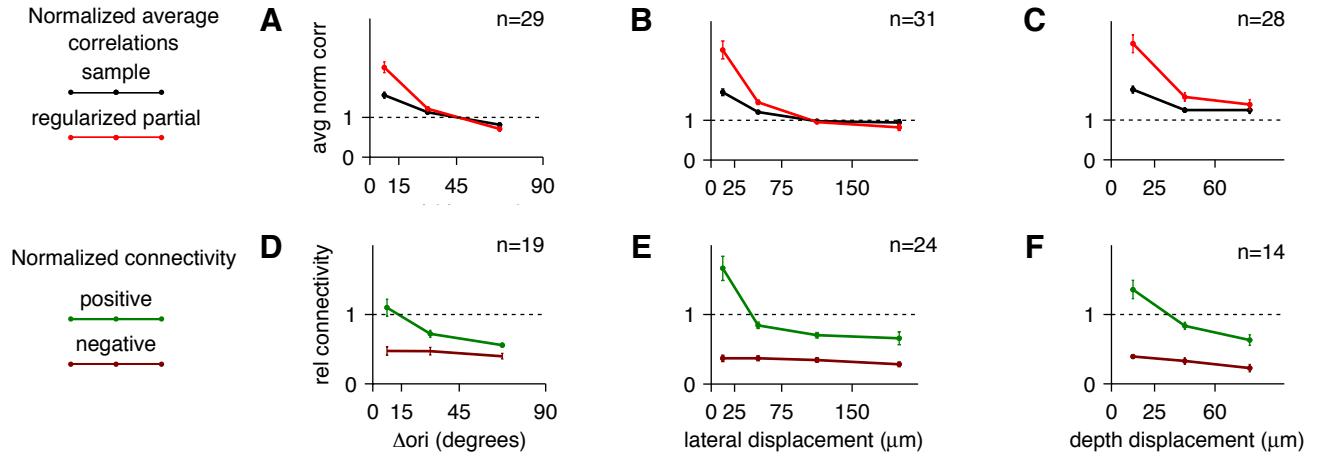


Figure 6. Dependence of sample correlations and regularized partial correlations on orientation tuning and physical distance. **A–C.** Average partial correlations (red) estimated by $C_{\text{sparse}+\text{latent}}$ and average sample correlations (black) averaged across multiple imaged sites. At each site the correlations were normalized by the respective average correlation shown in Fig. 5 D. The number of sites, n , that qualified to be included in the analysis is indicated in each panel. 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 based on the number of sites 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 $C_{\text{sparse}+\text{latent}}$. 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.

spatial arrangement and types of recorded cells. Recently, advanced imaging techniques have allowed for recordings from nearly every cell in a volume of cortical tissue *in vivo* [54,55] and even from entire nervous systems [69,70]. These techniques may provide more incisive measurements of functional connectivity than classical 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 [23,26,29,58]. 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 our 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 [71], we *a priori* favored solutions with sparse partial correlations. The importance of partial correlations may be justified by the principle of maximum entropy: The maximum entropy

distribution on discrete or continuous multivariate domains constrained by the observed mean firing rates, their variances, and correlations is the multivariate normal distribution, where the precision matrix specifies the interactions between cell pairs. Therefore, under the aforementioned assumptions that the recorded population is sufficiently complete and that the model correctly represents the nature of interactions, the network of partial correlations can be hypothesized to be a better representation of functional dependencies than correlations.

Functional connectivity as coactivations Another approach to describing the functional connectivity of a circuit is to isolate patterns of multineuronal coactivations [72–77]. 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 [72, 75], *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 [78].

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 through modes of coactivations was represented by the factor analysis 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 [79], high-order correlations [80], or global network fluctuations [81].

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 [49, 50]. 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 [82]. 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 and that the stationarity assumption is observed [81]. However, ambitious projects, such as the BRAIN initiative [47], 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 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. [80] developed a heuristic rule to identify the most significant features that must be fitted by a maximum entropy model for improved performance in retina. Generalized linear models typically employ L_1 penalty terms to constrain the solution space and to effectively reduce the dimensionality of the solution [61].

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

Model selection We evaluated the covariance matrix estimators using the cross-validated 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, would also serve as estimators of the covariance matrix. If a different model yields better estimation of the covariance matrix than the estimator proposed here, we believe that its structure should deserve consideration as the better representation of the functional connectivity.

The results of model selection must be interpreted with caution. As we demonstrated by simulation, with limited data, an incorrect but more constrained model can produce a better estimate than a more complex model with correct representation of dependencies. Therefore, showing that a more constrained model has better cross-validated performance than a more complex model does not necessarily support the conclusion that it reveals a better representation of dependencies in the data. This caveat is related to *Stein’s Paradox* [83]: The biasing of an estimate toward an arbitrary low-dimensional target is guaranteed to produce *some* improvement over the best estimate with many degrees of freedom.

Physiological interpretation and future directions Here 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 most significant sample correlations Fig. 4 F, H, and I. 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. Although the network of functional interactions inferred by the optimal estimator, $C_{\text{sparse+latent}}$, will not be identical to the anatomical connectivity, it may accurately reflect average connectivity profiles. Future experiments utilizing molecular markers for different cell types and follow-up multiple whole-cell *in vitro* experiments [23, 28] will 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 interpretation. 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).

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 [55]. Briefly, 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 the 3D-RAMP two-photon microscope as described in [55]. 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 sets of stimuli were presented for each imaging site: the first to map directional tuning and the second to estimate noise correlations. Directional tuning was mapped using a pseudo-random sequence of drifting gratings at sixteen equally spaced directions of motion, 500 ms per direction, for 3 min without blanks. The data for covariance estimation were collected during presentations of full-field drifting gratings with the same parameters as those used in directional tuning except only two directions (in 9 datasets) or five directions (in 22 datasets) were used and the presentations lasted 1 second and were separated by 1-second blanks. Each stimulus condition was presented between 100 and 300 times.

Data processing All data were processed in MATLAB using the DataJoint data processing chain toolbox (datajoint.github.com) first developed in our lab.

The collected fluorescent traces were deconvolved to reconstruct the firing rates for each neuron. First, the first principal component was subtracted from the traces, which reduced common mode noise related to small cardiovascular movements [55]. 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 [56].

Orientation tuning was computed by fitting the mean firing rates in response to gratings drifting in 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 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 -value of the fit was computed as the fraction of the permuted datasets for which the R^2 value of the tuning function fit exceeded that of the real data. Cells were considered tuned with p -values below 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 2% of the 90th percentile in each site) were exculded from the analysis.

Cross-validation To compare the performance of the estimators against each other, we used conventional 10-fold cross validation. 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 splitting of the data was chosen to achieve a high degree of independence between the 10 datasets for cross-validation purposes. Splitting the data into smaller blocks could increase dependencies between them due to temporal correlations thereby violating the assumptions of cross-validation. 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 the optimal hyperparameter values estimated within the inner loop. Hyperparameters were optimized by a two-phase search algorithm: random search to find a good starting point followed by pattern search to find the global minimum. The inner cross-validation loop subdivided the training dataset from the outer loop to perform 10-fold cross-validation in order to evaluate each choice of the hyperparameter values. Thus the size of the training dataset within the inner loop comprised 81% of the entire recording.

When the cross-validation loss was not required, only the inner loop of cross-validation was used, applied to the entire dataset. This approach was used to compute the covariance matrix estimates and

their excess-loss in the simulation study (Fig. 1 Rows 4 and 5) and to analyze the partial correlation structure of the sparse+latent estimator (Fig. 4–6).

Covariance estimation Within the inner loop of cross-validation, covariance matrix estimation was performed with fixed hyperparameter values provided by the search algorithm. The computation of regularized estimators 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:

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

Then the diagonal matrix D was used as the target of covariance shrinkage (Eq. 5) to produce the final regularized estimate. The `corpcor` package in the programming language R also implements this estimator [84], although its analytical optimization of the shrinkage intensities is based on the mean squared error whereas we optimized them with respect to the normal loss function (Eq. 9).

Estimator C_{factor} used two hyperparameters: the number of latent factors d and the shrinkage intensity $\lambda \in [0, 1]$. The matrix L of rank d and the diagonal matrix of individual variances Ψ were computed by solving the minimization problem

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

which we solved by an expectation-maximization (EM) algorithm. Under our chosen loss function (Eq. 9), this is equivalent to maximum likelihood estimation of L and 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 to regulate the sparsity of its precision matrix S . The precision matrix is computed in two steps: First, the zero structure is determined by minimizing the L_1 -penalized loss:

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

where $\tilde{Z} \succ 0$ denotes the constraint that \tilde{Z} be a positive definite matrix and $\|\tilde{Z}\|_1$ is the element-wise L_1 norm of the matrix \tilde{Z} . This problem formulation is known as *graphical lasso* [85, 86]. To solve this minimization problem, we modified the alternative-direction method of multipliers (ADMM) algorithm [50]. Then, after the zero structure was determined, the remaining coefficients were fitted without penalty:

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

where $Z^\#$ denotes the set of positive-definite matrices with the same zero structure as Z . 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 $C_{\text{sparse+latent}}$ has two hyperparameters: the number of latent units d and the sparsity level λ . $C_{\text{sparse+latent}}$ estimates a larger sparse precision matrix S^* of the joint distribution of the p observed neurons and d latent units.

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

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

population is

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

The $p \times p$ matrix $L = S_{12}S_{22}^{-1}S_{12}^T$ 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 the low-rank matrix L .

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 [49,50]:

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

Then we find the sparse and low-rank components of the precision matrix by solving (also using ADMM)

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

The partial correlation matrix with both sparse and low-rank components is computed from $C_{\text{sparse+latent}}$ according to Eq. 4; it includes the effects of interactions between the visible and latent units. This estimate of the partial correlations was used in analyses of average partial correlations (Fig. 4B and E and Fig. 5D, and Fig. 6A–C). The sparse component S was used separately in analyses of the connectivity graphs (Fig. 4F, G, H and Fig. 5B, C, E, and F and Fig. 6D–F); in this case, the partial correlations were computed by normalizing the sparse component alone without the low-rank component:

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

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

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

Acknowledgments

We thank Genevera Allen for a helpful discussion, and Eftychios Pnevmatikakis for helpful suggestions and feedback on the manuscript. This work was supported by grants NEI R01-EY018847, NEI P30-EY002520-33, NEI T32-EY07001-37, NIDA R01DA028525, the NIH-Pioneer award DP1-OD008301 to AST, the McKnight Scholar Award to AST; the Arnold & Beckman Foundation Young Investigator Award to AST; and NSF grants DMS-0817649, DMS-1122094 to KJ.

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. Ostožić 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. 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.

40. Peng J, Wang P, Zhou N, Zhu J (2009) Partial correlation estimation by joint sparse regression models. *Journal of the American Statistical Association* 104.
41. 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.
42. 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.
43. Ledoit O, Wolf M (2004) A well-conditioned estimator for large-dimensional covariance matrices. *Journal of multivariate analysis* 88: 365–411.
44. Bickel PJ, Li B, Tsybakov AB, van de Geer SA, Yu B, et al. (2006) Regularization in statistics. *Test* 15: 271–344.
45. 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.
46. Friedman JH (1989) Regularized discriminant analysis. *Journal of the American statistical association* 84: 165–175.
47. Alivisatos AP, Chun M, Church GM, Deisseroth K, Donoghue JP, et al. (2013) The brain activity map. *Science* 339: 1284–1285.
48. Dempster A (1972) Covariance selection. *Biometrics* : 157–175.
49. 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.
50. Ma S, Xue L, Zou H (2013) Alternating direction methods for latent variable gaussian graphical model selection. *Neural computation* : 1–27.
51. 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.
52. Fan J, Fan Y, Lv J (2008) High dimensional covariance matrix estimation using a factor model. *Journal of Econometrics* 147: 186–197.
53. Reddy GD, Saggau P (2005) Fast three-dimensional laser scanning scheme using acousto-optic deflectors. *J Biomed Opt* 10: 064038.
54. 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* .
55. Cotton RJ, Froudarakis E, Storer P, Saggau P, Tolias AS (2013) Three-dimensional mapping of microcircuit correlation structure. *Frontiers in Neural Circuits* 7: 151.
56. 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.
57. Feldt S, Bonifazi P, Cossart R (2011) Dissecting functional connectivity of neuronal microcircuits: experimental and theoretical insights. *Trends in neurosciences* 34: 225–236.

58. 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.
59. Alonso JM, Martinez LM (1998) Functional connectivity between simple cells and complex cells in cat striate cortex. *Nature neuroscience* 1: 395–403.
60. Aertsen A, Gerstein G, Habib M, Palm G (1989) Dynamics of neuronal firing correlation: modulation of “effective connectivity”. *Journal of neurophysiology* 61: 900–917.
61. 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.
62. 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.
63. Tkacik G, Schneidman E, Berry I, Michael J, Bialek W (2006) Ising models for networks of real neurons. *arXiv preprint q-bio/0611072* .
64. Yu S, Huang D, Singer W, Nikolic D (2008) A small world of neuronal synchrony. *Cereb Cortex* 18: 2891–901.
65. 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.
66. 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.
67. Zalesky A, Fornito A, Bullmore E (2012) On the use of correlation as a measure of network connectivity. *Neuroimage* 60: 2096–2106.
68. 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.
69. Leung LC, Wang GX, Mourrain P (2013) Imaging zebrafish neural circuitry from whole brain to synapse. *Frontiers in neural circuits* 7.
70. 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.
71. 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.
72. Gerstein GL, Bedenbaugh P, Aertsen AM (1989) Neuronal assemblies. *Biomedical Engineering, IEEE Transactions on* 36: 4–14.
73. Chapin JK, Nicolelis MA (1999) Principal component analysis of neuronal ensemble activity reveals multidimensional somatosensory representations. *Journal of neuroscience methods* 94: 121–140.
74. 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.
75. 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.

76. Lopes-dos Santos V, Conde-Ocazinez S, Nicolelis MA, Ribeiro ST, Tort AB (2011) Neuronal assembly detection and cell membership specification by principal component analysis. *PLoS one* 6: e20996.
77. Lopes-dos Santos V, Ribeiro S, Tort AB (2013) Detecting cell assemblies in large neuronal populations. *Journal of Neuroscience Methods* 220: 149 – 166.
78. 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.
79. 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*
- .
80. 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.
81. 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* .
82. 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.
83. Efron B, Morris CN (1977) Stein's paradox in statistics. WH Freeman.
84. Schäfer J, Opgen-Rhein R, Strimmer K (2010) corpcor: Efficient estimation of covariance and (partial) correlation. *R package version 15 7*.
85. Meinshausen N, Bühlmann P (2006) High-dimensional graphs and variable selection with the lasso. *The Annals of Statistics* 34: 1436–1462.
86. Friedman J, Hastie T, Tibshirani R (2008) Sparse inverse covariance estimation with the graphical lasso. *Biostatistics* 9: 432–441.

Supporting Information

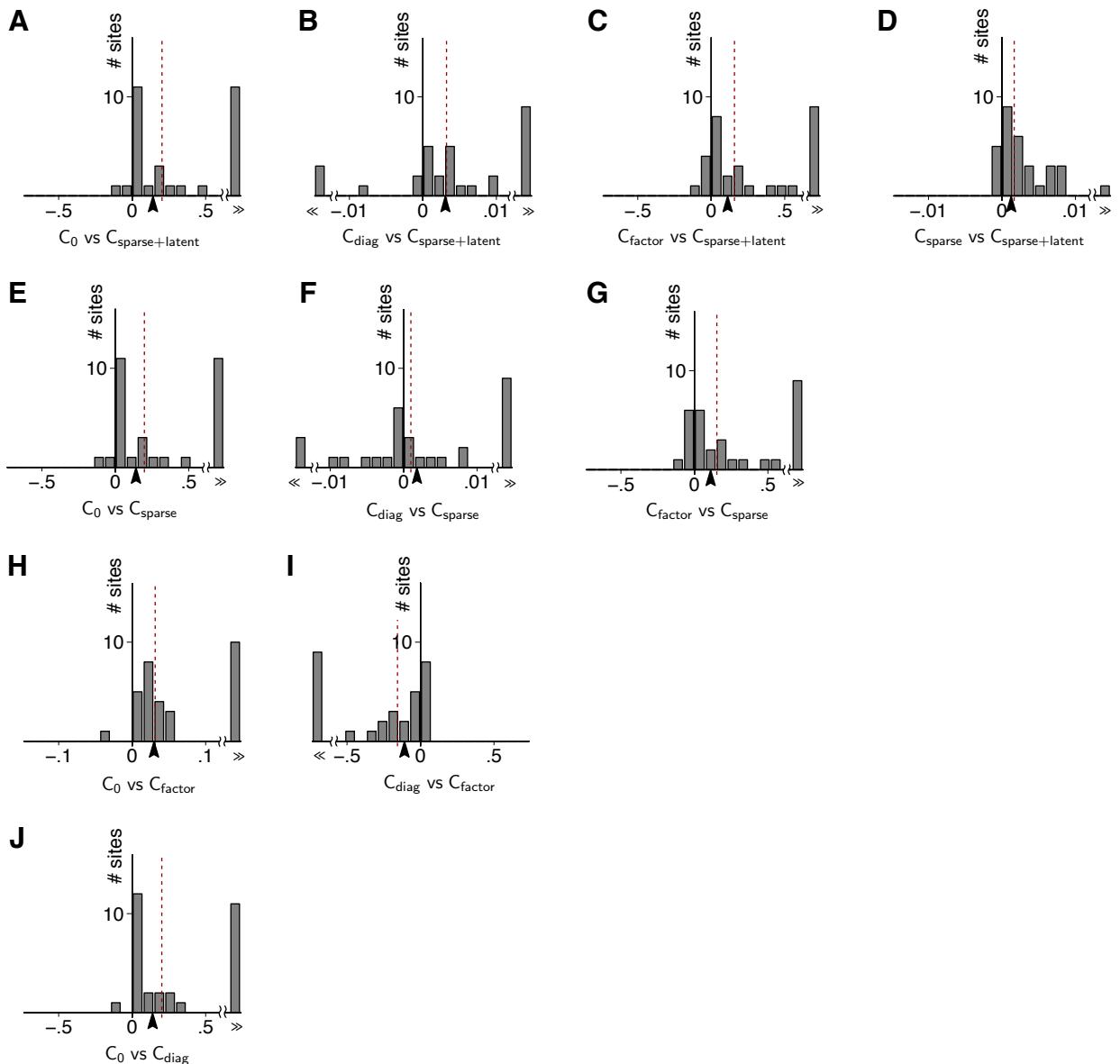


Figure S 1 (preceding page). All-to-all performance comparisons of the sample covariance matrix and the four regularized estimators with respect to multivariate normal cross-validation loss.

- A–D. These panels are identical to those in Fig. 3 A–D.
- E. C_{sparse} outperforms C_0 : median improvement 0.15 nats/neuron, $p = 3.9 \times 10^{-5}$.
- F. C_{sparse} performs similarly to C_{diag} : median improvement 1.0×10^{-3} nats/neuron, $p = 0.13$.
- G. C_{sparse} outperforms C_{factor} : median improvement 0.15 nats/neuron, $p = 1.2 \times 10^{-4}$.
- H. C_{factor} outperforms C_0 : median improvement 0.029 nats/neuron, $p = 2.7 \times 10^{-4}$.
- I. C_{factor} performs worse than C_{diag} : median improvement -0.16 nats/neuron, $p = 1.6 \times 10^{-4}$.
- J. C_{diag} outperforms C_0 : median improvement 0.14 nats/neuron, $p = 2.7 \times 10^{-5}$.

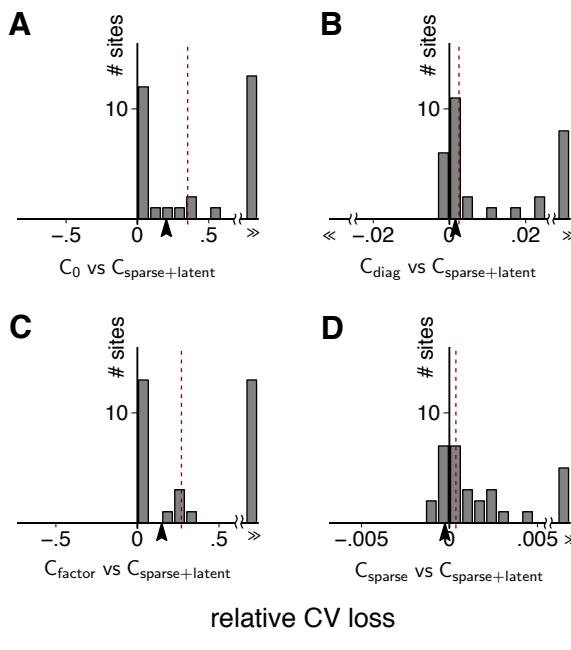


Figure S 2. The sparse+latent estimator outperforms the other estimators under the quadratic loss (Eq. 14) just as it did under the normal loss (Eq. 9) used in Fig. 3. Histograms of the relative cross-validation loss (Eq. 13) of estimators C_0 , C_{diag} , C_{factor} , and C_{sparse} with respect to $C_{\text{sparse+latent}}$ using the quadratic loss function. The histograms were obtained from 31 imaged sites in 15 mice. All medians (red dashed lines) were significantly greater than zero, indicating the dominance of $C_{\text{sparse+latent}}$ over the other estimators. The arrow heads indicate the results for the site shown in Fig. 2 and Fig. 4. The p -values are for the two-sided Wilcoxon's signed rank test.

- A. $C_{\text{sparse+latent}}$ outperforms C_0 : median $\Delta\ell_{C_0, C_{\text{sparse+latent}}} = 0.35$, $p = 1.7 \times 10^{-6}$.
- B. $C_{\text{sparse+latent}}$ outperforms C_{diag} : median $\Delta\ell_{C_{\text{diag}}, C_{\text{sparse+latent}}} = 2.5 \times 10^{-3}$, $p = 1.0 \times 10^{-4}$.
- C. $C_{\text{sparse+latent}}$ outperforms C_{factor} : median $\Delta\ell_{C_{\text{factor}}, C_{\text{sparse+latent}}} = 0.27$, $p = 1.2 \times 10^{-6}$.
- D. $C_{\text{sparse+latent}}$ outperforms C_{sparse} : median $\Delta\ell_{C_{\text{sparse}}, C_{\text{sparse+latent}}} = 3.7 \times 10^{-4}$, $p = 2.2 \times 10^{-3}$.